IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF DELAWARE

NIPPON SHINYAKU CO., LTD., Plaintiff, v. SAREPTA THERAPEUTICS, INC., Defendant.	C.A. No. 21-1015 (GBW) **VOLUME 1 (Part 2) (Exhibits 5-12)**
SAREPTA THERAPEUTICS, INC., Defendant and Counter-Plaintiff, v. NIPPON SHINYAKU CO., LTD. and NS PHARMA, INC. Plaintiff and Counter-Defendants.	

JOINT APPENDIX TO CLAIM CONSTRUCTION BRIEF
FOR THE WILTON/UWA PATENTS

MORGAN, LEWIS & BOCKIUS LLP
Amy M. Dudash (#5741)
1201 North Market Street, Suite 2201
Wilmington, DE 19801
(302) 574-3000
amy.dudash@morganlewis.com

Attorneys for Plaintiff/Counterclaim Defendant Nippon Shinyaku Co., Ltd. and Counterclaim Defendant NS Pharma, Inc.

MORRIS, NICHOLS, ARSHT & TUNNELL LLP
Jack B. Blumenfeld (#1014)
Megan E. Dellinger (#5739)
1201 North Market Street
P.O. Box 1347
Wilmington, DE 19899
(302) 658-9200
jblumenfeld@morrisnichols.com
mdellinger@morrisnichols.com

Attorneys for Defendant/Counter-Plaintiff Sarepta Therapeutics, Inc.

OF COUNSEL:

Amanda S. Williamson
Christopher J. Betti
Krista V. Venegas
Maria E. Doukas
Zachary Miller
Guylaine Haché
Michael T. Sikora
MORGAN, LEWIS & BOCKIUS LLP
110 North Wacker Drive, Suite 2800
Chicago, IL 60601
(312) 324-1000

Eric Kraeutler
MORGAN, LEWIS & BOCKIUS LLP
1701 Market Street
Philadelphia, PA 19103
(215) 693-5000

OF COUNSEL:

Charles E. Lipsey
J. Derek McCorquindale
Ryan O'Quinn
FINNEGAN, HENDERSON, FARABOW, GARRETT & DUNNER, LLP
1875 Explorer Street, Suite 800
Reston, VA 20190-6023
(571) 203-2700

William B. Raich
Michael J. Flibbert
Yoonhee Kim
Yoonjin Lee
FINNEGAN, HENDERSON, FARABOW, GARRETT & DUNNER, LLP
901 New York Avenue, NW
Washington, DC 20001-4413
(202) 408-4000

Alissa K. Lipton
FINNEGAN, HENDERSON, FARABOW, GARRETT & DUNNER, LLP
Two Seaport Lane
Boston, MA 02210-2001
(617) 646-1600

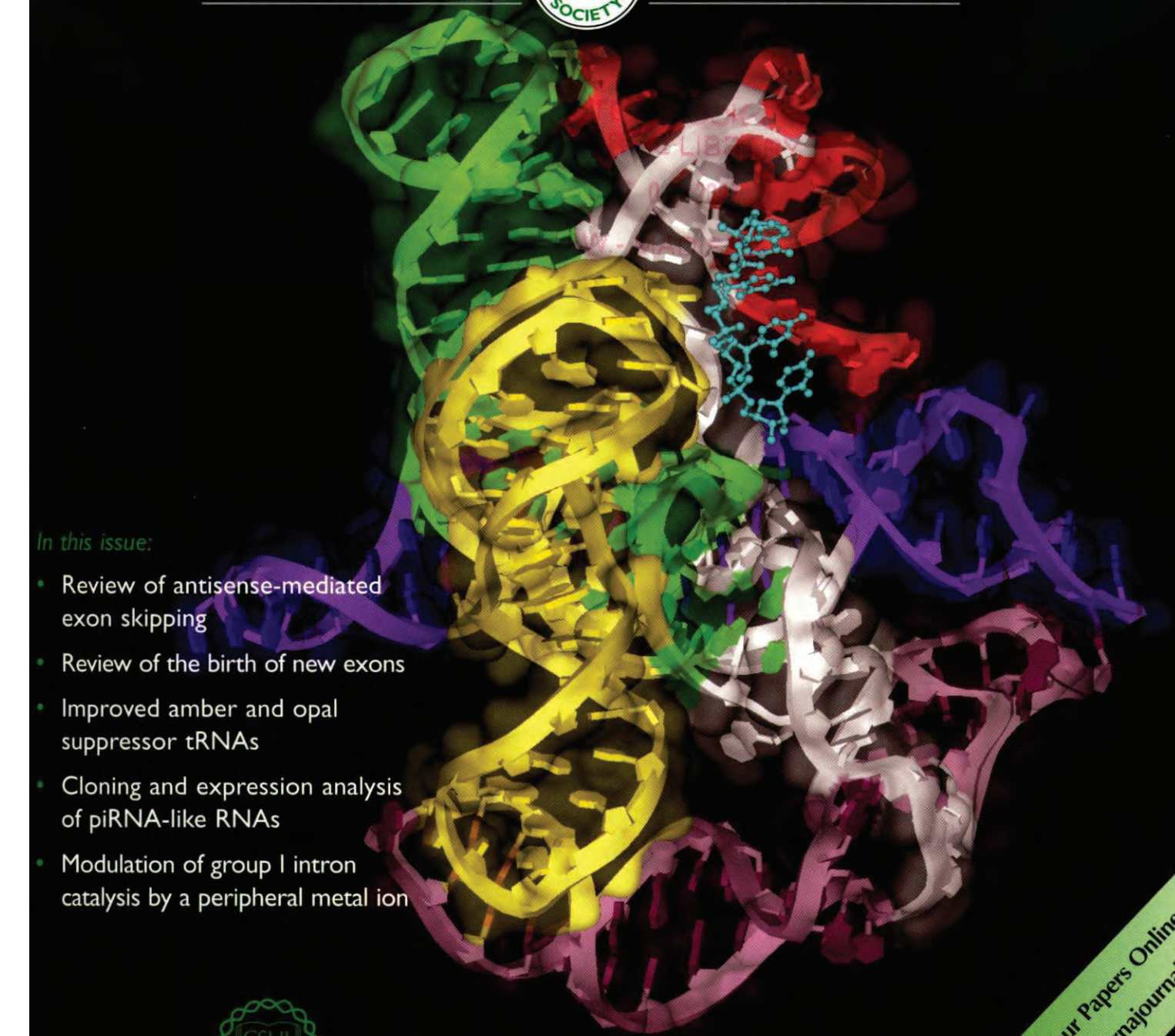
March 20, 2023

EXHIBIT 5

www.rnajournal.org

RNA

VOL. 13, NO. 10 OCTOBER 2007



In this issue:

- Review of antisense-mediated exon skipping
- Review of the birth of new exons
- Improved amber and opal suppressor tRNAs
- Cloning and expression analysis of piRNA-like RNAs
- Modulation of group I intron catalysis by a peripheral metal ion

COLD SPRING HARBOR LABORATORY PRESS

ISSN 1355-8382

RNA

A PUBLICATION OF THE RNA SOCIETY

Editorial Office: *RNA*, Center for RNA Molecular Biology, Room W127, School of Medicine, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106-4973, USA. Telephone: 216-368-1852; Fax: 216-368-2010; E-mail: rnajournal@case.edu.

Publishing and Advertising Offices: Cold Spring Harbor Laboratory Press, 1 Bungtown Road, Cold Spring Harbor, New York 11724-2203. For advertising rates and data, contact Marcie Siconolfi, Advertising Manager, at 516-422-4010 (telephone) or 516-422-4092 (fax).

Instructions for Authors: Instructions for authors are published in each issue of *RNA* and are available on the World Wide Web (http://www.rnajournal.org).

Books for Review: Publications should be sent to Timothy Nilsen, Center for RNA Molecular Biology, Room W127, School of Medicine, Case Western Reserve University, Cleveland, OH 44106-4973, USA. Readers are encouraged to draw to his attention publications that may be of interest to other readers of *RNA*.

Subscriptions: Published monthly. Subscribers have a choice of "online only" or "print + online" subscriptions for this journal. For 2007 institutional pricing, visit www.cshlpress.com/journal_info.tpl. Personal rate (must be paid by check, credit card, or money order): Online only: $220; Print + Online: U.S. $280, Canada and Mexico $340, R.O.W. $365. Contact Kathleen Cirone, Subscription Manager, Fulfillment Department, 500 Sunnyside Boulevard, Woodbury, New York 11797-2924. Telephone: Continental U.S. and Canada 1-800-843-4388; all other locations 516-422-4100. Fax: 516-422-4097. Claims for missing issues must be received within four months of issue date.

Member Subscriptions: Members of the RNA Society automatically receive *RNA* as part of their annual membership dues.

The RNA Society: Membership information and application forms may be obtained from: The RNA Society, 9650 Rockville Pike, Bethesda, MD 20814-3998, USA. Telephone: 301-530-7120; Fax: 301-530-7049; E-mail: rna@faseb.org.

RNA (ISSN 1355-8382) is published monthly by Cold Spring Harbor Laboratory Press, 500 Sunnyside Blvd., Woodbury, NY 11797-2924. Periodicals paid at Woodbury, NY, and additional mailing offices. **POSTMASTER:** Send address changes to Cold Spring Harbor Laboratory Press, Journals Fulfillment Dept., 500 Sunnyside Blvd., Woodbury, NY 11797-2924.

RNA
A PUBLICATION OF THE RNA SOCIETY
VOL. 13, NO. 10 OCTOBER 2007

CONTENTS

(*continued*)

Cover Illustration: Crystal structure of a phage Twort group I ribozyme-product complex (PDB code: 1y0q; Golden, B.L., Kim, H., and Chase, E. 2005. Crystal structure of a phage Twort group I ribozyme product complex. Nat. Struct. Mol. Biol. ***12:*** *82–89). Image details: ribozyme derived from the second group I intron in the* orf142 gene (orf142-I2)*: ribbon–plate representation, transparent surface, P1-P2 domain—red, P3-P7 region—green, P4-P6 domain—blue, P9-P9.1 domain—purple, P7.1-P7.2 subdomain—yellow, nucleotides not included in these domains—white; oligonucleotide representing a 5′ exon: ball-and-stick representation, cyan. The image was generated with the Accelrys Discovery Studio Visualizer. Cover image provided by the Jena Library of Biological Macromolecules (JenaLib; www.fli-leibniz.de/IMAGE.html).*

Published by Cold Spring Harbor Laboratory Press

Methods

OA Open Access paper

REVIEW

Antisense-mediated exon skipping: A versatile tool with therapeutic and research applications

ANNEMIEKE AARTSMA-RUS and GERT-JAN B. VAN OMMEN

DMD genetic therapy group, Department of Human Genetics, Leiden University Medical Center, 2300 RC, Leiden, The Netherlands

ABSTRACT

Antisense-mediated modulation of splicing is one of the few fields where antisense oligonucleotides (AONs) have been able to live up to their expectations. In this approach, AONs are implemented to restore cryptic splicing, to change levels of alternatively spliced genes, or, in case of Duchenne muscular dystrophy (DMD), to skip an exon in order to restore a disrupted reading frame. The latter allows the generation of internally deleted, but largely functional, dystrophin proteins and would convert a severe DMD into a milder Becker muscular dystrophy phenotype. In fact, exon skipping is currently one of the most promising therapeutic tools for DMD, and a successful first-in-man trial has recently been completed. In this review the applicability of exon skipping for DMD and other diseases is described. For DMD AONs have been designed for numerous exons, which has given us insight into their mode of action, splicing in general, and splicing of the *DMD* gene in particular. In addition, retrospective analysis resulted in guidelines for AON design for *DMD* and most likely other genes as well. This knowledge allows us to optimize therapeutic exon skipping, but also opens up a range of other applications for the exon skipping approach.

Keywords: exon skipping; splicing; Duchenne muscular dystrophy; antisense oligonucleotides; therapy

INTRODUCTION

Antisense oligonucleotides (AONs) are mostly known for their ability to hybridize to a sense target sequence, which leads to RNase H cleaving of the RNA:DNA hybrid and results in specific gene expression knockdown (Hausen and Stein 1970; Zamecnik and Stephenson 1978). This approach offered useful opportunities to study development because it allowed timed gene knockdown in early or later stages of development, as well as therapeutic opportunities to knockdown genes involved in cancer, inflammatory diseases, and viral infections. Currently, an AON to treat CMV-induced retinitis (Vitravene) has been registered as a drug, and other AONs to treat cancer and inflammatory diseases are in phase II and III clinical trails (Marwick 1998; Kurreck 2003). However, with the emergence of RNAi, which turned out to be a more efficient and more predictable tool for expression knockdown, the field of AON-induced knockdown has gone in decline (Elbashir et al. 2001). A notable exception is the modulation of pre-mRNA splicing to induce exon skipping, where RNase H-independent AONs are employed to block splicing signals (Kole and Sazani 2001). This approach has gained increasing interest over the past decade (van Deutekom and van Ommen 2003). Actually, antisense-mediated exon skipping is currently one of the most promising therapeutic approaches for Duchenne muscular dystrophy (DMD). A first-in-man trial has recently been completed successfully in our institute (J.C.T. van Deutekom, A.A.M. Janson, I.B. Ginjaar, W.S. Frankhuizen, A. Aartsma-Rus, M. Bremmwe-Bout, J.T. den Dunnen, K. Koop, A.J. van der Kooi, N.M. Goemans, et al., in prep.) and a second trial is about to start in the United Kingdom (Muntoni et al. 2005; F. Muntoni, pers. comm.). This review describes the mechanism of antisense-mediated exon skipping for DMD and gives an overview of other exon skipping applications reported thus far. It discusses how the numerous AONs designed for DMD exon skipping give us insight into splicing of the *DMD* gene in particular, but splicing in general as well. Finally, ways to implement exon skipping in future applications will be discussed.

Reprint requests to: Annemieke Aartsma-Rus, DMD genetic therapy group, Department of Human Genetics, Leiden University Medical Center, P.O. Box 9600, 2300 RC, Leiden, The Netherlands; email: a.m.rus@lumc.nl; fax: 31-71-5268285.

Article published online ahead of print. Article and publication date are at http://www.rnajournal.org/cgi/doi/10.1261/rna.653607.

DUCHENNE MUSCULAR DYSTROPHY

Duchenne muscular dystrophy is a severely invalidating, progressive neuromuscular disorder (Emery 2002). Patients

are wheelchair bound before the age of twelve, often require assisted ventilation later in life, and generally die in their early twenties. The disease is caused by mutations in the *DMD* gene that abolish the production of functional dystrophin (Hoffman et al. 1987). This protein consists of two essential functional domains connected by a central rod domain that is partly dispensable (Hoffman et al. 1988; Koenig et al. 1988). Dystrophin links the cytoskeleton to the extracellular matrix and is thought to be required to maintain muscle fiber stability during contraction (Matsumura and Campbell 1994). Mutations that disrupt the open reading frame result in prematurely truncated proteins unable to fulfill their bridge function (Fig. 1). Ultimately, this leads to muscle fiber damage and the continuous loss of muscle fibers, replacement of muscle tissue by fat and fibrotic tissue, impaired muscle function, and eventually the severe phenotype observed for DMD patients.

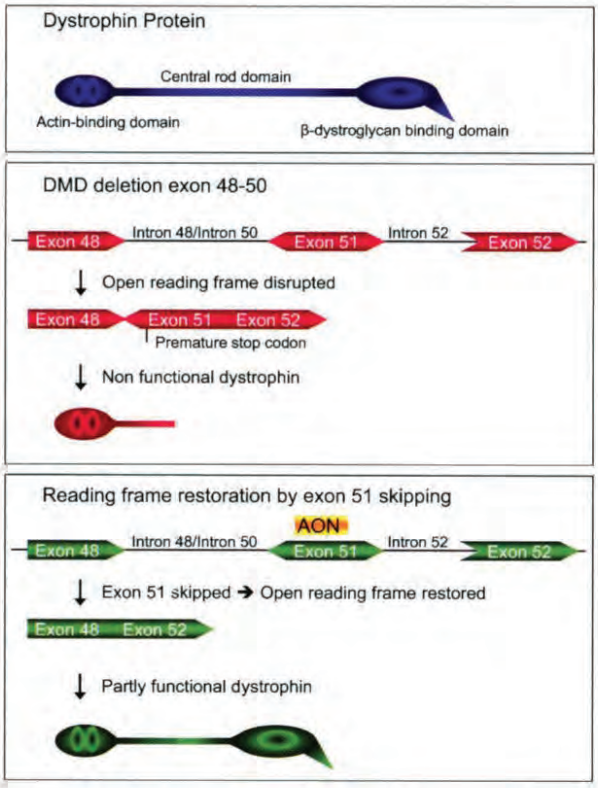


FIGURE 1. Antisense-mediated exon skipping for Duchenne muscular dystrophy. The dystrophin protein (*upper* panel) contains an N-terminal actin-binding domain connected to a β-dystroglycan binding domain by the central rod domain. Dystroglycan is a transmembrane protein that is bound to the extracellular protein laminin-2. Dystrophin thus fulfills a bridge function in muscle fibers by linking the cytoskeletal actin to the extracellular matrix. In Duchenne muscular dystrophy (*middle* panel), the open reading frame is disrupted (in this example by a deletion of exons 48–50, the most common mutation in DMD patients), resulting in a premature stop codon and a truncated dystrophin, which is unable to fulfill its bridge function. Antisense oligoribonucleotides (AONs) can be employed to restore the open reading frame (*lower* panel). Specific AONs hybridize to exon 51 and hide this exon from the splicing machinery, resulting in the splicing of exon 51 with its flanking intron. This restores the open reading frame, allowing the generation of an internally deleted dystrophin, that contains both the actin- and dystroglycan binding domains and therefore is partially to largely functional.

In contrast, mutations that maintain the open reading frame allow for the generation of internally deleted, but partially functional dystrophins (Monaco et al. 1988). These mutations are associated with Becker muscular dystrophy (BMD), a much milder disease when compared to DMD. Patients generally remain ambulant until later in life and have near normal expectancies, although more severely affected patients have been reported as well (Emery 2002).

The *DMD* gene is the largest known human gene and its 79 exons span an astonishing 2.4 Mb (Roberts et al. 1993, 1994). Over 70% of all DMD and BMD patients suffer from deletions of one or multiple exons (Aartsma-Rus et al. 2006c). Mildly affected BMD patients carrying deletions that involve over two thirds of the central rod domain have been described, suggesting that this domain is largely dispensable. Dystrophin can be largely functional as long as the N- and C-terminal domains are present to convey the link between the cytoskeleton and the extracellular matrix (England et al. 1990; Mirabella et al. 1998). Due to this rather unique feature, the counterintuitive skipping of additional internal exons can be employed to enlarge a deletion, but at the same time restore the open reading frame and thus convert a severe DMD into a milder BMD phenotype (Fig. 1; van Deutekom et al. 2001).

THE EXON SKIPPING APPROACH

Antisense-mediated modulation of pre-mRNA splicing has been pioneered by Ryszard Kole (Dominski and Kole 1993). In the first experiments, AONs were aimed at activated cryptic splice sites in the β-globin (*HBB*) and cystic fibrosis transmembrane conductance regulator (*CFTR*) genes in order to restore normal splicing in β-thalassemia and cystic fibrosis patients (Dominski and Kole 1993; Sierakowska et al. 1996; Friedman et al. 1999). Even though this approach does not technically qualify as exon skipping (but rather the redirection of normal splicing), it does offer therapeutic potential for diseases where mutations often induce cryptic splice sites such as the Hutchinson–Gilford progeria syndrome (Scaffidi and Misteli 2005). In fact, for most genetic disorders an estimated 5%–10% of mutations induce abnormal splicing (Krawczak et al. 1992; Cartegni et al. 2002), part of which can, in principle, be corrected.

A finding of the group of Matsuo eventually alerted the DMD field to a potential therapeutic application of exon skipping for DMD. Matsuo and colleagues observed that a 52-base pair (bp) deletion within exon 19 resulted in the skipping of this exon in the so-called DMD Kobe patient (Matsuo et al. 1990, 1991). This hinted at the presence of a motif within this 52-bp deletion required for proper inclusion of exon 19 in the mRNA. Indeed, AONs targeting

part of this deletion induced exon 19 skipping in vitro and in human control lymphoblastoma cells (Takeshima et al. 1995; Pramono et al. 1996).

The feasibility of the approach was then studied in parallel in patient-derived cell lines and in cells from the *mdx* mouse model. This mouse carries a nonsense point mutation in the in-frame exon 23 (Sicinski et al. 1989). Thus, by skipping exon 23 the nonsense mutation is bypassed while the reading frame is maintained. Proof of principle on RNA level was obtained first in cultured muscle cells from the *mdx* mouse by two groups independently (Dunckley et al. 1998; Wilton et al. 1999). In both cases, the reading frame was restored on RNA level as analyzed by RT-PCR analysis. Our group was the first to show restoration of dystrophin on protein level after targeted exon 46 skipping in cultured muscle cells from two DMD patients with an exon 45 deletion (van Deutekom et al. 2001). The wide therapeutic applicability was then confirmed by others and us in numerous patient-derived cell cultures (Takeshima et al. 2001; Aartsma-Rus et al. 2003, 2004a; Surono et al. 2004; Aartsma-Rus et al. 2007). The majority of these mutations involved deletions of one or more exons, but reading frame restoration for nonsense point mutations and single exon duplications has been reported as well. Notably, for single exon duplications, skipping either one of the duplicated exons will restore the wild-type transcript and dystrophin protein (Aartsma-Rus et al. 2007). Some mutations require the skipping of two exons in order to restore the reading frame. We confirmed that this so-called double exon skipping is indeed feasible using a combination of individual AONs targeting the two different exons (Aartsma-Rus et al. 2004a). Remarkably, the efficiency of this double exon skipping approach was only slightly lower than that of single exon skipping (~70%–75% versus 75%–80% dystrophin-positive myotubes, respectively). In parallel, results in *mdx* mouse underlined the therapeutic promise of exon skipping (Mann et al. 2001, 2002; Lu et al. 2003). Local intramuscular injections of an optimized AON resulted in ~20% of wild-type dystrophin levels accompanied by improvement in muscle histology and function (Lu et al. 2003). Dystrophin protein was detectable by Western blot analysis for at least 3 months after a single intramuscular injection.

In theory, exon skipping would be applicable to the majority of DMD patients. Exceptions are mutations located between exon 64 and exon 70, which are essential for protein function, deletions that abolish all actin-binding sites in the N-terminal region or involve the first or the last exon, and large chromosomal rearrangements such as translocations. These mutations are uncommon and make up less than 10% of all mutations (Aartsma-Rus et al. 2006c). Thus exon skipping can theoretically be applicable for up to 90% of DMD patients (Aartsma-Rus et al. 2004a).

A disadvantage of the AON approach is that it is mutation specific in that different mutations require the skipping of different exons to restore the open reading frame. Fortunately, DMD deletions and duplications mainly occur in two hot spot regions, i.e., the major hot spot region (involving exon 45 to exon 53) and the minor hot spot region (located between exon 2 and exon 20) (Liechti-Gallati et al. 1989; Beggs et al. 1990; White et al. 2006). Therefore, by strategically choosing target exons, through the skipping of eight different exons, this strategy would be therapeutic for over 50% of all patients (van Deutekom and van Ommen 2003; Aartsma-Rus and van Deutekom 2007). The most notable example is exon 51 skipping, which is applicable to almost 25% of DMD patients with a deletion, or 16% of all DMD patients (Aartsma-Rus and van Deutekom 2007).

To obtain proof of concept in humans, a "first-in-man study" on exon skipping was undertaken by our center in collaboration with Prosensa B.V. using 2′-*O*-methyl phosphorothioate AONs (chemistries will be discussed in more detail later) (J.C.T. van Deutekom, A.A.M. Janson, I.B. Ginjaar, W.S. Frankhuizen, A. Aartsma-Rus, M. Bremmwe-Bout, J.T. den Dunnen, K. Koop, A.J. van der Kooi, N.M. Goemans, et al., in prep.). Four DMD patients received a single, local intramuscular injection with AONs targeting exon 51 and a biopsy was taken one month later. Preliminary results are very promising and no serious adverse effects were observed or reported by the patients as a result of AON injection. Another local study using morpholino AONs is to start soon in the United Kingdom. These first-in-man studies are an important step toward the clinical application of antisense-mediated exon skipping for DMD.

A systemic pilot study has been performed by Takeshima and colleagues in a single DMD patient at a very low dosage using phosphorothioate RNA (0.5 mg/kg) (Takeshima et al. 2006).

AON DESIGN AND MODE OF ACTION

The first targets to induce exon skipping are the donor and acceptor splice sites and the branch point sequence. These sites have indeed been successfully targeted in the majority of the exon skip applications, including exon skipping for DMD (Table 1; Dunckley et al. 1998; Mann et al. 2002; Wilton and Fletcher 2005). However, they consist of consensus sequences shared with many different genes and consequently targeting them involves the risk of mistargeting splice sites of other genes. Alternatively, it has now been shown that exon skipping can be induced by targeting exon-internal sites, which has been successful in the *DMD* and *WT1* genes (van Deutekom et al. 2001; Renshaw et al. 2004; Aartsma-Rus et al. 2005; Wilton and Fletcher 2005). Proper recognition by the splicing machinery and inclusion into the mRNA is thought to depend on exonic splicing enhancer (ESE) motifs for the majority of exons (Cartegni et al. 2002). These sites are involved in exon recognition through the binding of members of a subfamily of splicing

TABLE 1. Overview of exon skipping applications

Target gene	Protein	Target[a]	Goal	Application	Reference[b]
APOB	Apolipoprotein B	3′ SS and BP exon 27	Knockdown of APOB100 isoform	Retard atherosclerosis	Khoo et al. (2007)
Bcl-X	Bcl-xS and Bcl-xL	5′ SS Bcl-xL exon	Isoform switching from anti- to pro-apototic Bcl-x	Cancer therapy	Mercatante et al. (2001, 2002)
COL7A1	Collagen type 7	EI Exon 70	Allele specific knockdown	Dystrophic epidermolysis bullosa therapy	Goto et al. (2006)
DMD	Dystrophin	3′ SS, 5′ SS, EI numerous DMD exons	Reading frame restoration leading to partially functional dystrophins	DMD therapy	van Deutekom et al. (2001) Aartsma-Rus et al. (2003) Lu et al. (2003) Aartsma-Rus et al. (2004a) Alter et al. (2006)
FOLH1	Prostate-specific membrane antigen	5′ SS exon 1, exon 6, or exon 18	Isoform switching from transmembrane to cytoplasmatic form	Prostate cancer therapy	Williams and Kole (2006)
IL-5Ralpha	IL-5 receptor-α	3′ SS or 5′ SS exon 9	Isoform switching from transmembrane to soluble form	Asthma therapy	Karras et al. (2000, 2001)
MyD88	MyD88	5′ SS exon 2	Isoform switching	Anti-inflammatory	Vickers et al. (2006)
Tau	Tau	5′ SS or 3; SS exon 10	Restore normal ratio 3R/4R tau isoform	FTDP-17[c] therapy	Kalbfuss et al. (2001)
TNFRSF1B	TNFα 2 receptor	Exon 7 and 8	Isoform switching from transmembrane to soluble form	Rheumatoid arthritis therapy	P. Sazani (pers. comm.)
Ttn	Titin	5′ SS exon 45, 79, 37, 47	Isoform specific knockdown	Functional analysis of isoforms	Seeley et al. (2007)
WT1	WT1	IE Exon 5	Isoform switching to pro-apoptotic form	Leukemia therapy	Renshaw et al. (2004)

[a]SS, splice site; IE, intra-exonic; BP, branch point site.
[b]An overview of the most important publications for each application.
[c]Frontotemporal dementia and parkinsonism linked to chromosome 17.

factors, known as serine and arginine rich proteins (SR proteins) (Stojdl and Bell 1999). These SR proteins have one or several RNA domains able to bind to loosely defined sequence motifs that make up ESEs. SR proteins then recruit the essential U2AF and U1 snRNP splicing factors to the 3′ polypyrimidine tract and 5′ splice sites, respectively, and thus facilitate splicing. The importance of ESEs is underlined by the finding that intraexonic point mutations often result in exon skipping on the RNA level, rather than yielding no or missense amino acid changes as deduced from DNA analysis (Cartegni et al. 2002). Famous examples are the *neurofibromatosis type 1* gene and the *ataxia telangiectasia mutated* gene, where a significant number of mutations lead to exon skipping (Teraoka et al. 1999; Ars et al. 2000; Wimmer et al. 2007). In addition, predicted nonsense mutations in in-frame exons of the DMD gene occasionally turn out to actually induce exon skipping and a BMD phenotype, indicative that these nonsense mutations disrupt ESE sites (Shiga et al. 1997; Ginjaar et al. 2000; Tuffery-Giraud et al. 2004; Disset et al. 2006). As SR protein binding to ESEs is essential for exon inclusion, blocking ESEs with AONs would be expected to result in exon skipping. Matsuo and colleagues indeed showed that blocking the ESE they had identified in exon 19 resulted in exon skipping (Pramono et al. 1996). ESE motifs are only loosely defined because, even though inclusion of the exon in mRNA is essential, strict motifs would interfere with the main task of an exon, i.e., to encode protein information. Therefore, targeting ESEs reduces the chance of mistargeting. Software packages, such as RESCUE-ESE, ESEfinder, and the PESX server, predict putative ESE sites (Fairbrother et al. 2002; Cartegni et al. 2003; Zhang and Chasin 2004; Smith et al. 2006), which facilitates the design of exon-internal AONs. We have now designed almost 150 exon-internal AONs, of which nearly 70% are effective in inducing the skipping of 39 different DMD exons (2, 8, 17, 19, 29, 33, 40–64, 71–78) (A. Aartsma-Rus and J.C.T. van Deutekom, unpubl.; Aartsma-Rus et al. 2005). Initially,

not much was known about ESE sites besides that they were thought to be purine rich, probably due to the fact that the most abundant SR proteins, SF2/ASF and SC35, can bind these motifs (Tacke and Manley 1995). We reasoned that, in order to be able to be bound by SR proteins, ESE sites had to be open regions in the secondary RNA structure. Therefore, our initial 114 exon-internal were mainly directed against purine-rich sequences located in open regions in the secondary RNA structure as predicted by the Mfold server (Zuker 2003). We found that two out of three (78 out of 114) of these AONs were effective in inducing the skipping of 36 exons in control muscle cell cultures (Aartsma-Rus et al. 2005).

No difference was observed between the "openness" of the target sequence in the predicted secondary structure of effective versus ineffective AONs (Aartsma-Rus et al. 2005). Because of its extreme size it was difficult to predict the secondary structure of the entire DMD gene or even of a target exon with its flanking introns before splicing. Thus, AON design was based on the secondary structure of an exon and 100 bp of flanking intron sequences. Analysis of larger regions revealed that often the local secondary structure of the exon and its immediate surroundings were present within the larger secondary structure. For each exon there were numerous predicted structures that were often more or less equally energetically stable. Our AON design was based on the most likely secondary structure. However, it is likely that a certain pre-mRNA exists in more than one secondary structure in the nucleus. Therefore, it may be better to analyze the number of times each nucleotide in the target sequence is present in an open structure in all predicted structures or, in other words, the propensity of the nucleotide to be single stranded (SS) in each of the predicted structures (also known as the SS count) (Zuker 2003). Nevertheless, when we calculated this propensity for target sequences of effective and ineffective AONs, no significant differences were observed (M. Hirshi and A. Aartsma-Rus, unpubl.). At first sight, this would suggest that the level of openness is irrelevant for AON efficacy. However, as our AONs were designed to target open structures in the first place, this finding is biased. In addition, the longest stretch of nucleotides predicted never to be single stranded in any of the predicted structures was eight, and our AONs were 17–21 nucleotides (nt) long. This made it hard to verify whether AONs targeting completely closed structures are indeed ineffective. The closest one can come to a successful comparison is by noting that our empirical approach of selecting partly open structures has a 2 out of 3 success rate, while in several other approaches typically more AONs need to be designed to achieve proper exon skipping (Mann et al. 2002).

The availability of ESE predicting software allowed retrospective analysis of our set of 114 AONs for the presence or absence of putative ESE sites. Interestingly, compared to ineffective AONs, effective AONs targeted significantly more RESCUE ESE hexamers and significantly higher values for SF2/ASF- and SC35-binding sites as predicted by ESEfinder v2.0 (Aartsma-Rus et al. 2005, 2006a). When we compared the highest value for any of the four SR proteins for which ESEfinder has an algorithm (i.e., the most likely ESE), the difference between effective and ineffective AONs became even more significant (Aartsma-Rus et al. 2005). This suggests that exon-internal AONs indeed act by steric blocking of SR protein binding. This finding was further underlined by the fact that effective AONs were located significantly closer to the acceptor splice sites (Aartsma-Rus et al. 2005), and ESEs located within 70 nt of the acceptor splice sites have been reported to be more active than ESEs beyond this distance (Wu and Maniatis 1993; Fairbrother et al. 2004). Using ESE-predicting software to fine tune our AON design improved our success rate from 70% to ~75% (A. Aartsma-Rus, unpubl.). On comparison, GC content and AON length were similar for effective and ineffective AONs (Aartsma-Rus et al. 2005). Nevertheless it was recently reported that, for some AONs, increasing the length enhanced exon skipping efficiency (Harding et al. 2007). Our own recent studies indicate that the efficiency of some AONs that induce very low levels of exon skipping can be enhanced by increasing AON length, whereas increasing the length of an already efficient AON did not enhance AON efficiency and occasionally even reduced exon skipping levels (H.A. Heemskerk, G.-J.B. van Ommen, S. de Kimpe, P. van Kuik, J.C.T. van Deutekom, and A. Aartsma-Rus, in prep.). This phenomenon was also occasionally observed by Harding and colleagues (Harding et al. 2007).

Recently, the Wilton group described AON sequences to induce the skipping of each DMD exon except the first and last (Wilton et al. 2007). AON design was based on targeting the donor or acceptor splice sites or putative ESE sites as predicted by ESEfinder v3.0 (Smith et al. 2006). AONs that proved effective were then further optimized by designing overlapping AONs specific for the same target site; 470 AONs were screened in this manner. Of the optimal AON for each exon, 56 target exon-internal sequences, 16 the 3′ (acceptor) splice sites, and one the 5′ (donor) splice site. For 26 of the 56 exons skipped by exon-internal AONs, we had already identified effective AONs (Aartsma-Rus et al. 2005). In 25 out of 26 cases the optimal target sequences overlapped partially or completely with the target sites we determined earlier. Interestingly, the percentage of exon-internal AONs is highest in the most efficiently skipped group (83%), lower in the medium efficient group (73%), and lowest (64%) in the group of exons that can barely be skipped. The finding that exon-internal AONs appear to be more efficient than blocking the splice sites themselves contrasts with the fact that mutations abolishing a splice site result in exon skipping for virtually all cases, while mutations abolishing an ESE site often lead to partial exon skipping (Ginjaar et al. 2000;

Deburgrave et al. 2007). A possible explanation is that the splice site motifs are more sharply defined, while ESEs are more of a pattern. Thus, a single mutation in an ESE will have a less pronounced effect than one affecting a splice site. In contrast, binding of U1 snRNP and U2 snRNP to the donor and acceptor splice sites, respectively, may occur with a higher affinity than SR protein binding to ESE sites. This would imply that U1 and U2 may be better able to compete with AON binding than SR proteins. Moreover, as SR proteins recognize motifs rather than consensus sequences, it is not unlikely that the binding site is partly determined by a specific secondary structure. AON binding to ESEs is likely to disrupt the local secondary structure, and would so further prevent SR protein binding.

Given the enormous length of DMD introns, it is likely that DMD exons are more dependent on ESEs for recognition by the splicing machinery than exons of other genes. Therefore, we hypothesized that some exons might have two or more mutually exclusive ESE sites (Aartsma-Rus et al. 2006b). This would render them insensitive to steric hindrance of SR protein binding to one of those sites, while blocking both should induce exon skipping. We indeed recently reported efficient and reproducible skipping of three thus far unskippable exons (exons 47, 57, and 64) and of the poorly skippable exon 45 using a combination of exon-internal AONs (Aartsma-Rus et al. 2006b). The potential of double targeting was confirmed by Wilton and colleagues, who reported that for exons 10, 20, 34, and 65, which were poorly or not skippable with individual AONs, skipping at high levels could be induced using a mix of two or even three (exon 65) AONs (Wilton et al. 2007).

AONS TO STUDY SPLICING

Regardless of whether exon-internal or splice site AONs are used, the aim of antisense-mediated exon skipping has thus far been to disrupt the splicing of the targeted exon. To determine if exon characteristics such as exon length, length of the flanking introns, and/or strength of the predicted splice sites affected the levels of exon skipping (or the "skippability" of an exon), we compared said characteristics for exons that could be skipped at either high, medium, or low levels as reported by Wilton and colleagues (Wilton et al. 2007). We observed that the predicted acceptor splice sites of poorly skippable exons were significantly higher than those of exons that could be skipped at medium or high efficiency ($P = 0.04$, Kruskal–Wallis signed rank sum test) (Fig. 2). It makes sense that exons with poorly defined splice sites, which critically depend on ESEs, are easier to skip. This finding explains why for exons, which are skipped at low levels, the percentage of AONs targeting the acceptor splice sites was much higher than for efficiently skipped exons (83% versus 64%). This implies that when one is free to choose a target exon within a transcript for antisense-mediated exon skipping, it is probably best to choose exons with low predicted values for acceptor splice sites. Interestingly, no significant difference was observed for the strengths of donor splice sites between the different groups (Fig. 2). Combined with the finding that for only one single exon out of the 77 tested the donor splice site was the optimal target, this must imply that donor splice sites are of lesser importance for exon definition during DMD splicing. This is in contradiction with the current view on exon definition, which states that exons are defined by binding of splicing factors first to the donor and then to the acceptor splice site of an exon (Robberson et al. 1990). It is possible that this view needs revisiting, or that DMD splicing is atypically complex, with its intron sizes varying from 107 bp to 248 kb, thus DMD exons may not behave as other exons. Indeed, for several other genes, the donor (5′) splice site has been targeted successfully (Table 1).

No relation was observed between exon skipping efficiency and exon length, or lengths of the upstream, downstream, or flanking introns (Fig. 2). In general, only the targeted exon was skipped, but occasionally unexpected results were observed. A striking example is that AONs targeting exon 8 always induce skipping of both exon 8 and 9 in human and dog transcripts (Aartsma-Rus et al. 2005; McClorey et al. 2006b; Wilton et al. 2007). The most likely explanation for this phenomenon is that the AONs somehow affect only the acceptor splice site but leave the donor splice site intact. Therefore, exon 8 can be joined to exon 9 and both exons are then spliced out together due to the AONs that disrupt the acceptor splice site of exon 8 (Fig. 3). This is consistent with the similar finding that an AON targeting the acceptor splice site of exon 17 resulted in the skipping of both this and the subsequent exon, whereas AONs targeting intraexonic sequences (abolishing donor splice site recognition) only cause single exon 17 skipping. However, AONs targeting acceptor splice sites can also result in skipping of only the targeted exon. Generally, upon the disruption of an acceptor splice site, the preceding donor splice site will be joined to the first accessible downstream acceptor splice site, leading to skipping of the targeted exon only. Thus, other factors such as the order of intron splicing must be involved as well. Transcription of the DMD gene takes ~16 h and the RNA has been shown to be cotranscriptionally spliced (Tennyson et al. 1995). However, it is unlikely that splicing occurs consecutively for all introns. In the case of exon 8, the preceding intron 7 is a hundred times longer than the following intron 8 (110 versus 1.1 kb), and splicing of the long intron 7 may take much more time so that in most transcripts exons 8 and 9 will already be joined before intron 7 is spliced out. This will enable the skipping of both exons 8 and 9 when AONs are used that do not disrupt the donor splice site of exon 8. We note that since it is not always possible to foresee whether AONs will induce

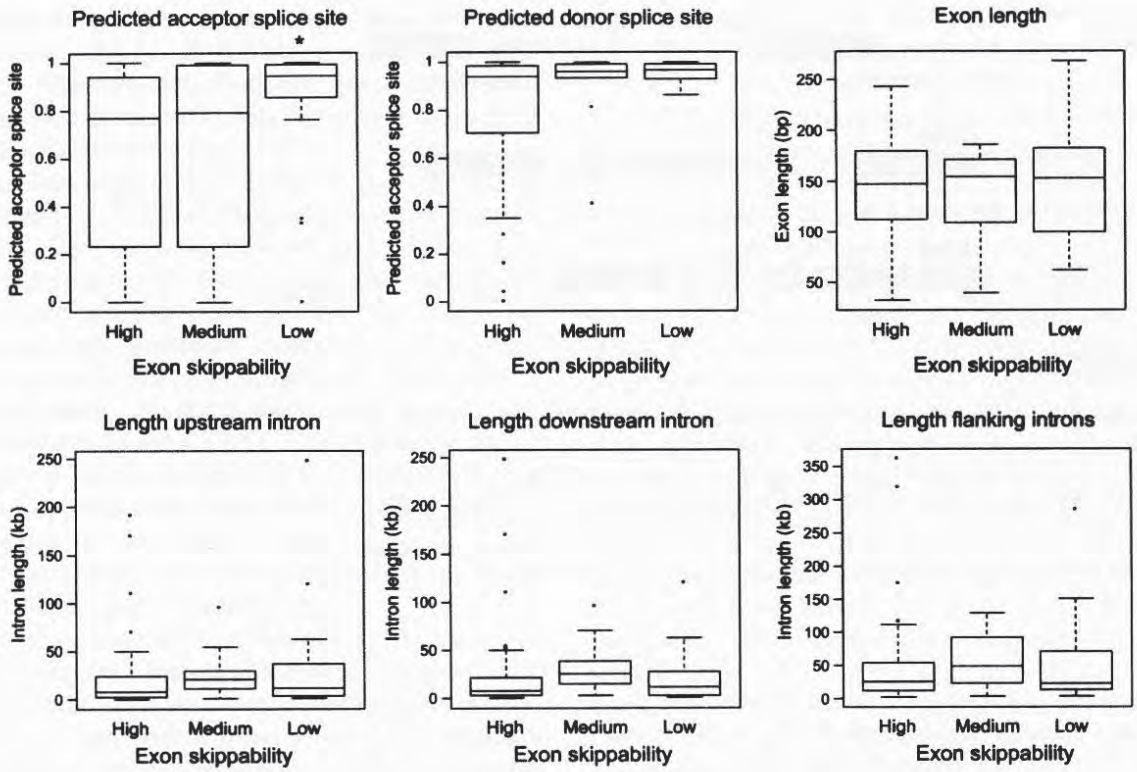


FIGURE 2. Box plots of the different groups of exons for the predicted donor and acceptor splice sites, exon length, and the lengths of the upstream, downstream, or flanking introns. Exon skippability is based on the report published by Wilton and colleagues (Wilton et al. 2007). The median value is indicated by a broad vertical line which is located within a box that contains all values between the 25th and 75th percentiles. The outer ranges are depicted by dotted lines and bordered by small horizontal lines. Outlying values are indicated by small circles. Splice site values were calculated with the Berkeley *Drosophila* Genome Project software for human splice site prediction. The predicted acceptor splice sites were significantly higher for poor skippable exons as calculated with the Kruskal–Wallis signed rank sum test (*P*-value 0.04, indicated with an asterisk). No significant differences were observed for the other parameters, although there is a trend for predicted donor splice sites to be somewhat lower for highly skippable exons (*P*-value 0.2).

skipping of a single or multiple exons, it is advisable to analyze not just the RNA directly flanking the targeted RNA, as this may lead to misinterpretation of results. For example, using primers in exons 7 and 9 would have led us to believe the exon 8 AONs were ineffective, because no skipping would have been observed. Another example is exon 23 skipping in the *mdx* mouse, where single exon 23 skipping is often accompanied by the out-of-frame skipping of exons 22 and 23 (Mann et al. 2001, 2002). In addition, using primers further away, the occasional skipping of longer stretches of exons has been reported (Dunckley et al. 1998; Fall et al. 2006). AONs targeting exon 54 induced equal levels of single exon 54 skipping and skipping of both exons 54 and 55, suggesting that intron 54 splicing occasionally precedes intron 53 splicing. These results imply that AONs can be a tool to study the splicing process per se in more detail. The timing and sequence of intron splicing is likely dependent on intron length, but nucleotide composition and secondary RNA structure may play a role as well. Knowing the sequence of intron splicing of a certain gene can explain the outcome of splicing mutations that result in complex splicing patterns (Schwarze et al. 1999).

The nonconsecutive splicing of the DMD gene explains why it is sometimes feasible to skip multiple consecutive exons targeting only the two outer exons (so-called multi-exon skipping). We first observed multiexon skipping after treating patient and control muscle cultures with AONs specific for exons 45 and 51, in a successful attempt to correct the reading frame of an exon 46–50 deletion (Aartsma-Rus et al. 2004a). In the control myotubes we observed low levels of single and double exon 45 and 51 skipping, but also exon 45–51 skipping. Notably, exon 45–51 skipping would be therapeutic for 13% of all DMD patients and would thus reduce the mutation specificity of the exon skipping approach. Furthermore, it would allow the generation of larger deletions, e.g., those known to be associated with a milder BMD phenotype. Unfortunately, despite many attempts it proved unfeasible to induce multiexon skipping of exons 17–48/51 and 48–59, which would be large deletions associated with extremely mild phenotypes (Aartsma-Rus et al. 2006b). In retrospect, this is not surprising given the long transcription time and the cotranscriptional splicing (Tennyson et al. 1995). Thus, exon 16 will be joined to exon 17 long before exon 48 is even transcribed (an estimated 4.5 h

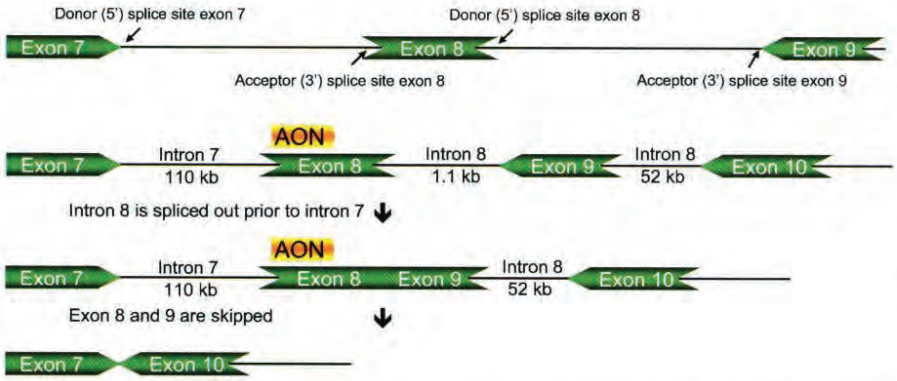


FIGURE 3. Antisense-mediated exon 8 skipping. Using AONs targeting exon 8 only the skipping of both exons 8 and 9 is observed. A likely explanation is that splicing of the downstream intron 8 (1.1 kb) precedes splicing of intron 7 (110 kb) and that the AONs do not affect the donor splice site of exon 8 (effective AONs used so far target either the acceptor splice site or the 5′ region of the exon). Thus, intron 8 can be spliced out and exons 8 and 9 are joined. As exon 8 AONs do disrupt the acceptor splice site of exon 8, the splicing machinery uses the first available acceptor splice site, which is that of exon 10 (because exons 8 and 9 are already joined).

later). Therefore, in order to induce multiexon skipping, one may need to target each individual exon in this stretch. This has been achieved at relatively high efficiencies in *mdx* mouse in vitro and in vivo for up to seven exons (Fall et al. 2006). However, targeting more exons was less efficient and typically led to many intermediate products where some, but not all, of the intended exons are skipped (S. Wilton, pers. comm.). Exon 42–55 and 45–60 skipping proved feasible but not consistently, and we often observed exon 45–55 skipping both in treated and nontreated samples (Aartsma-Rus et al. 2006b). Exon 45–55 skipping would potentially be therapeutic for ~40% of all DMD patients and an exon 45–55 deletion has been observed in asymptomatic individuals (Beroud et al. 2007). This multiexon skipping is probably more feasible than the very large skips, as the stretch involves less exons, while the flanking introns (44 and 55) are very long (248 kb and 120 kb, respectively). Thus, it is likely that for a significant number of transcripts exons 45–55 may already be joined, before introns 44 and 55 are spliced out, making exon 45–55 multiexon skipping an attractive target. Our preliminary results indicate that exon 45–55 multiexon skipping targeting the outer exons 45 and 55 is feasible, but levels are as yet too low to be beneficial (A. Aartsma-Rus, L. van Vliet, J.C.T. Deutekom, and G.-J.B. van Ommen, unpubl.).

EXON SKIPPING FOR OTHER APPLICATIONS

Reading frame restoration

The applicability of the exon skipping approach is not restricted to DMD (for an overview, see Table 1). A limited number of proteins share the feature of dystrophin that an in-frame, internal deletion is compatible with partially functionality of proteins. One example is type VII collagen. Truncating mutations in the *COL7A1* gene are associated with dystrophic epidermolysis bullosa, a disease characterized by severe blistering of the skin (Uitto et al. 1995; Fine et al. 2000). In contrast, mutations leading to in-frame exon skipping result in milder cases, suggesting that reading frame restoration might have therapeutic potential for this disease (McGrath et al. 1999). Using mutation-specific AONs targeting exon 70 resulted primarily in the skipping of the mutated exon, while the normal exon was included in the mRNA. Type VII collagen lacking the 16 amino acids encoded by exon 70 showed near normal functionality (Goto et al. 2006). Over 20% of recessive dystrophic epidermolysis bullosa patients carry a mutation in exon 70, making this approach a promising therapeutic tool for a significant subset of patients.

Isoform switching

Alternatively, AONs have been used to change levels of alternatively spliced genes. A striking example is the *Bcl-x* gene, which has two isoforms, Bcl-xS and Bcl-xL, that arise from two different 5′ splice sites in exon 2 (Mercatante et al. 2001, 2002). The Bcl-xS isoform is pro-apoptotic and sensitizes cells to chemotherapy. In contrast, Bcl-xL is anti-apoptotic and induces resistance to chemotherapeutic agents. Using AONs targeting the xL splice site, it was possible to shift the alternative splicing patterns toward the xS isoform. In vitro this shift by itself resulted in massive apoptosis in some cancer cell lines, and each cell line tested became sensitive to several chemotherapeutic agents (Mercatante et al. 2002). Unfortunately, after systemic delivery of AONs, one of the main target organs is the liver. Bcl-x AONs indeed induced liver apoptosis in treated mice after tail vein injection (Williams and Kole 2006), thus limiting the applicability of this approach.

A gene that is often inappropriately overexpressed in leukemia and solid tumors is the Wilms' tumor gene (*WT1*) (Scharnhorst et al. 2001). The gene product is thought to interfere with normal signaling, leading to maintenance of a malignant phenotype by increased proliferation and inhibition of differentiation and apoptosis. Most leukemic cells express high levels of *WT1* transcripts containing the alternatively spliced exon 5 (Renshaw et al. 2004). AON-mediated exon 5 skipping led to loss of cell viability and a decrease in cell survival in leukemic cell cultures (Renshaw et al. 2004). The expression of the *WT1*

gene in adults is restricted to specific cell types in kidney, gonads, hematopoietic cells, the nervous system, and mesothelium (Reddy and Licht 1996). In vivo analysis of the AONs will have to determine potential adverse effects of exon 5 skipping in these tissues.

Another anti-cancer approach described is the use of AONs to induce isoform switching of the prostate-specific membrane antigen, encoded by the *folate hydrolase* gene (Williams and Kole 2006). This protein is mainly expressed in prostate cells and one isoform is 140-fold higher expressed in malignant versus normal prostate tissues (O'Keefe et al. 2004). The overexpressed isoform has a functional enzymatic domain that is located extracellularly and regulates folate uptake (Davis et al. 2005). Other prostate-specific membrane antigen isoforms exist, which arise from the alternative splicing of exons 6 and 18 (Williams and Kole 2006). In these isoforms the enzymatic domain is not present or inactive, respectively. A fourth isoform results from an alternative donor splice site in the first exon (Su et al. 1995). This isoform lacks the transmembrane domain of the antigen and as a consequence the enzyme activity is sequestered in the cytoplasm. Individual AONs targeting exon 1, 6, and 18 were able to induce isoform switching, which was accompanied by lower levels of the full-length isoform and decreased enzymatic activity (Williams and Kole 2006). AONs targeting the donor splice site of the full-length first exon are the most promising for therapeutic application, as the isoform without the membrane domain may be involved in a pro-apoptotic pathway (Williams and Kole 2006). In addition, side effects are expected to be low as in adults the expression of the full-length isoform is restricted to malignant cells.

Frontotemporal dementia and parkinsonism linked to chromosome 17 (FTDP-17) results from tau deposits in the frontotemporal lobe of brain and neuronal cell death (Spillantini and Goedert 1998). Tau is normally involved in microtubule assembly and stability and has either three or four microtubule-binding domains (3R and 4R isoforms, respectively) (Goedert et al. 1989). The ratio between these isoforms is tightly regulated, as indicated by the finding that the majority of FTDP-17 patients carry mutations in the alternatively spliced exon 10 that lead to enhanced inclusion of this exon and increased expression of the 4R isoform (Varani et al. 1999). Using AONs targeting exon 10, levels of the 4R tau isoform were decreased in cultured rat cells that normally predominantly express the 4R isoform (Kalbfuss et al. 2001). This was accompanied by a changed morphology of the cytoskeleton of treated cells, suggesting that these AONs can be used to study the function of the different isoforms. It will be challenging to make this strategy therapeutically applicable, as the ratio between the 3R and 4R isoforms is tightly regulated and the target tissue (brain) is difficult to reach.

Isoform shifting is also useful for numerous inflammatory diseases. Receptors involved in immune stimulatory signal transduction, such as the IL-5 or TNFα receptors, also have an isoform that lacks the transmembrane domain. The soluble receptors bind to their ligand without induction of the inflammatory pathways and thus neutralize the bioactivity of the ligand. In many inflammatory diseases an effector has been identified, e.g., IL-5 for asthma and TNFα for rheumatoid arthritis (Panayi et al. 2001; O'Byrne 2006). Therefore, antisense-mediated isoform shifting has therapeutic potential for these diseases. AONs targeting exon(s) coding for the transmembrane induced efficient exon skipping, resulting in both a decrease of the membrane bound isoform and an increase in the soluble, neutralizing receptor (Karras et al. 2000, 2001; P. Sazani, pers. comm.).

An alternative approach to inhibit inflammation is modulation of the splicing of MyD88, which is an adaptor protein involved in IL-1β-dependent NFκB activation (Burns et al. 1998). $MyD88_L$ is translated from a full-length transcript, whereas the $MyD88_S$ isoform arises from the alternative exclusion of exon 2. $MyD88_L$ binds the IL-1 receptor and IRAK-1 (Interleukin receptor-associated kinase 1) and recruits IRAK-4, leading to the phosphorilation of IRAK-1 and subsequent NFκB activation. $MyD88_S$ is unable to recruit IRAK-4 and is, therefore, unable to induce NFκB activation (Burns et al. 2003). As $MyD88_S$ acts in a dominant-negative way, isoform switching from $MyD88_L$ to $MyD88_S$ may have therapeutic potential to treat inflammatory diseases associated with excessive IL-1 receptor signaling, such as atherosclerosis. AONs targeting exon 2 of the *MyD88* gene were indeed able to induce isoform switching, which was accompanied by diminishing pro-inflammatory signaling through the IL-1 receptor in vitro and in vivo (Vickers et al. 2006).

Gene knockdown studies

Recently, exon skipping has also been applied as an alternative way to achieve gene knockdown. This has advantages when compared to the standard RNase H gene knockdown as exemplified by AON-mediated exon skipping of Apolipoprotein B (*APOB*). There are two natural APOB isoforms. The full-length APOB100 protein is required for the assembly of VLDL, IDL, and LDL, is one of the ligands for the LDL receptor, and plays a central role in atherosclerosis (Soutar and Naoumova 2007). The other isoform, APOB48, is essential for chylomicron assembly and intestinal fat transport (Chester et al. 2000). APOB48 arises from intestine tissue-specific RNA editing of a CAA into a UAA termination codon in exon 26 (Chester et al. 2000). As a consequence, APOB48 lacks the LDL receptor binding domain. APOB100 knockdown is under investigation as a potential treatment for atherosclerosis. However, RNAi- and RNase H-induced degradation will result in knockdown of the detrimental APOB100 and the essential APOB48 isoform. Khoo and colleagues used the antisense-mediated exon skipping approach to target exon 27 of the

APOB transcript (Khoo et al. 2007). As the RNA editing signal for *APOB48* is located in exon 26, exon 27 skipping will not affect this isoform. It will, however, disrupt the open reading frame of the *APOB100* transcript, leading to lower amounts of APOB100 and likely to lower LDL and cholesterol levels. This hypothesis is backed up by the finding that heterozygote individuals with truncating mutations in exon 27 have low LDL and cholesterol levels and are resistant to the development of atherosclerosis (Whitfield et al. 2004). Exon 27-specific AONs indeed resulted in skipping of the targeted exon leading to a truncated, nonfunctional APOB100 protein, while APOB48 levels were maintained (Khoo et al. 2007).

The exon skipping approach has finally been employed to achieve isoform specific knockdown in order to determine isoform functionality. The human titin gene consists of 363 exons, and over a hundred alternatively spliced transcripts have been described (Freiburg et al. 2000). The isoforms are categorized into long N2A and N2B isoforms and the short Novex isoforms. Titin is proposed to serve as a template for sarcomere assembly, but as yet little is known about the function of the different titin domains and isoforms. In zebrafish, there are two titin orthologs, *ttna* and *ttnb*, which are highly homologous and both can give rise to N2A and N2B isoforms, whereas only *ttna* can encode the Novex isoform (Seeley et al. 2007). Using AONs it was feasible to induce skipping in a homolog and isoform-specific way, allowing the dissection and analysis of the function of the different titin isoforms in zebrafish development (Seeley et al. 2007). This approach showed that different titin isoforms have distinct functions, e.g., the N2A domain is required for sarcomere assembly in the somites, while both the N2A and N2B domains are essential for sarcomere assembly in the heart.

AONs to induce exon inclusion

AONs have also been used to induce inclusion of exons that are skipped due to mutations that disrupt ESEs (Cartegni and Krainer 2003; Hua et al. 2007). The best-studied example is the survival of motor neuron 2 (*SMN2*) gene, which is a homolog of the *SMN1* gene that is mutated in patients with spinal muscular atrophy (SMA) (Munsat and Davies 1992). *SMN2* is a near perfect homolog of *SMN1* but cannot compensate for the lack of SMN1 protein due to a translationally silent mutation in exon 7 of *SMN2*, which disrupts an SF2/ASF-binding site and results in an exon 7 skipping exon in most transcripts (Cartegni and Krainer 2002). As the amount of full-length *SMN2* transcripts is inversely correlated with disease severity, enhancing exon 7 inclusion is a putative therapy for SMA (Jablonka et al. 2000). Exon 7-specific AONs with a tail containing an ESE motif, or AONs linked to a serine-arginine peptide domain to recruit SF2/ASF to the disrupted ESE, resulted in higher levels of full-length SMN2 (Cartegni and Krainer 2003; Skordis et al. 2003). The same result was obtained with AONs targeting exonic splicing silencer motifs (Hua et al. 2007). Exonic splicing silencers are the counterparts of ESEs and are involved in the induction of exon skipping in, e.g., alternatively spliced exons. SMN protein levels increased after treatment with AONs targeting exonic splicing silencers, implying that AONs do not interfere with mRNA translation. Thus, this approach has therapeutic potential for SMA and possible for other diseases caused by mutations that disrupt ESEs or induce exonic splicing silencers.

TOWARD CLINICAL APPLICATIONS

The current review shows that antisense-mediated exon skipping is a promising tool for many research and therapeutic applications. If AONs manage to reach the cytoplasm, they will be effectively transported to the nucleus through a so far undefined mechanism. However, the main obstacle toward clinical application of this approach is the actual AON delivery to the target tissues. Biodistribution studies have shown that after systemic delivery, the majority of the AONs end up in the liver and the kidney for each of the different AON chemistries (Sazani et al. 2002; Fluiter et al. 2003; C.L. de Winter, H.A. Heemskerk, S. de Kimpe, P. van Kuik, G. Platenburg, and J.C.T. Deutekom, in prep.). On one hand, this is good news when the target gene is mainly and/or highly expressed in the liver, as is the case for, e.g., *APOB*. On the other hand, when the target gene is expressed in another tissue and in liver and kidney as well, AONs may trigger unwanted side effects, such as the liver apoptosis observed after treatment with Bcl-x AONs (Williams and Kole 2006). To obtain high local AON levels, direct injection into a tumor may be an option for some cancers, but injecting each and every muscle in DMD patients is unfeasible as muscle makes up 30% of the body and some muscles such as the diaphragm and the heart are difficult to reach. However, sometimes diseased tissues may also be more accessible; e.g., dystrophic muscle fibers are more permeable than healthy muscle fibers, resulting in enhanced intramuscular AON levels after systemic treatment (C.L. de Winter, H.A. Heemskerk, S. de Kimpe, P. van Kuik, G. Platenburg, and J.C.T. Deutekom, in prep.).

AON chemistry

Currently, different AON backbone chemistries are available, each having different characteristics. The most commonly used AON chemistry for splicing modulation is 2′-*O*-methyl or 2′-*O*-methoxyethyl RNA with a phosphorothioate (PS) backbone (Kurreck 2003). The 2′-*O*- modification renders the AON RNase H resistant and increases affinity for target RNA. The phosphorothioate backbone enhances stability as it inhibits AON breakdown by endo- and exonucleases. This modification is relatively cheap and

can be scaled up easily. Possible alternatives are morpholinos and locked nucleic acids (LNAs), which are both RNase H resistant. Morpholinos contain a six-membered morpholine moiety instead of the sugar ribose and phosphorodiamidate linkages (Summerton and Weller 1997). They have a nonionic backbone at physiological pH, making them notoriously hard to transfect in tissue culture experiments (Amantana and Iversen 2005). However, in vivo their nonionic nature results in higher tissue concentrations, due to the lack of nonspecific interactions with cellular components (Amantana and Iversen 2005). Morpholinos are nontoxic and very stable and have been shown efficient modulators of pre-mRNA splicing (Gebski et al. 2003). LNAs contain a methylene bridge that connects the 2′-*O* to the 4′-*C* of the ribose, forcing the nucleotide in the 3′ endoconformation (Obika et al. 1998). As a consequence LNAs are inflexible and have an extremely high affinity for RNA and DNA. In addition, they are nontoxic and nuclease resistant (Wahlestedt et al. 2000). LNAs have been reported to be extremely efficient modulators of pre-mRNA splicing (Aartsma-Rus et al. 2004b; Roberts et al. 2006). Ethylene-bridged nucleic acids (ENA) have an ethylene bridge instead of a methylene bridge and have comparable characteristics to LNAs (Morita et al. 2002, 2003).

To directly compare the effect of different AON analogs, the Kole group has developed an elegant read out system (Sazani et al. 2002). They generated a construct that contains a cryptic splice site in β-globin intron 2 linked to a green fluorescent protein gene. Without AONs the cryptic splice site will be used and GFP will not be produced, while effective AONs will redirect splicing and restore GFP synthesis. The amount of GFP reflects the efficiency of the AON. A mouse model stably expressing the GFP construct has been generated, allowing easy comparison of the biodistribution of different AON analogs (Sazani et al. 2002). Using this model, it was discovered that full-length LNAs generate an effect mainly in liver, colon, and small intestine after systemic delivery, thus providing a tool to manipulate splicing in these specific tissues (Roberts et al. 2006). Morpholinos have gained attention for DMD exon skipping since this chemistry is taken up at higher levels by the muscle (Sazani et al. 2002). Systemic delivery of morpholino in the *mdx* mouse was indeed more efficient than the 2′-*O*-methyl phosphorothioate counterpart (Alter et al. 2006; Fletcher et al. 2006; H.A. Heemskerk, G.-J.B. van Ommen, S. de Kimpe, P. van Kuik, J.C.T. van Deutekom, and A. Aartsma-Rus, in prep.). However, it is likely that the optimal chemistry partly depends on the target sequences as well. For example, LNAs were the most efficient chemistry to induce exon 46 skipping, but 2′-*O*-methyl PS AONs were optimal for exon 51 (A. Aartsma-Rus, unpubl.; Aartsma-Rus et al. 2004b). In addition, the morpholino targeting *mdx* mouse exon 23 is more efficient than the 2′-*O*-methyl counterpart, whereas for other exons morpholinos are equally efficient (McClorey et al. 2006a; H.A. Heemskerk, G.-J.B. van Ommen, S. de Kimpe, P. van Kuik, J.C.T. van Deutekom, and A. Aartsma-Rus, in prep.).

AON specificity

As the effect of the AONs is transient, due to AON clearance and breakdown of the targeted mRNA transcript and protein, patients will have to be treated repeatedly and chronically for genetic disorders like DMD. Therefore, one of the most important features of the AON has to be specificity, in order to avoid long-term side effects. During AON design BLAST analysis is performed for each AON to exclude annealing to other targets. However, this is based on the assumption that the AON only anneals to a completely homologous sequence. We compared the sequence specificity of LNAs, morpholinos, and 2′-*O*-methyl PS AONs (Aartsma-Rus et al. 2004b; H.A. Heemskerk, G.-J.B. van Ommen, S. de Kimpe, P. van Kuik, J.C.T. van Deutekom, and A. Aartsma-Rus, in prep.). As expected, LNAs showed very poor sequence specificity, and an AON containing two mismatches in a 14 mer was equally as efficient as the original LNA AON. The 2′-*O*-methyl PS AONs were much more sensitive to mismatches; a single mismatch either decreased exon skipping levels drastically or completely abolished AON efficacy. For morpholinos we obtained mixed results (H.A. Heemskerk, G.-J.B. van Ommen, S. de Kimpe, P. van Kuik, J.C.T. van Deutekom, and A. Aartsma-Rus, in prep.). In one case, two mismatches rendered the morpholino inefficient, for two other target sequences the same amount of mismatches reduced exon skipping levels or resulted in similar skipping levels as observed with the nonmismatched counterpart. In contrast, the mismatched 2′-*O*-methyl PS counterparts never induced a specific exon skipping.

AON delivery

As the majority of AON is cleared by the liver and kidneys, ways to decrease liver and kidney uptake and/or enhance tissue-specific uptake are under investigation. Certain cell-penetrating peptides enhance the uptake of morpholino AONs by muscle (Fletcher et al. 2007). This may also be achieved by linking muscle-homing molecules or peptides to AONs (Samoylova and Smith 1999; Ghosh and Barry 2005; Kolonin et al. 2006). Among the same line, one can envisage linking molecules or peptides specific for any given tissue to AONs to enhance tissue-specific uptake. Notably, alternative dystrophin isoforms are expressed in other tissues, such as the retina and the central nervous system. In DMD patients the muscle phenotype is most prominent, but deletions in the hot spot affect other isoforms as well. A complicating factor here is that the effect of the absence of the dystrophin isoforms in nonmuscle tissues is poorly understood. Nevertheless, AONs would

restore the reading frame of these isoforms as well. Thus for DMD the AONs do not necessarily have to target only muscle tissue. Actually, the fact that exon skipping strategy targets the pre-mRNA transcribed from the endogenous gene allows for the systemic restoration of different defective isoforms in a patient with one single AON.

Alternatively, the antisense sequence can be delivered to cells using viral vectors carrying a gene from which the antisense sequence can be transcribed, such as the small nuclear ribonucleoproteins (snRNPs). U7 snRNP is normally involved in histone processing and hybridizes to the spacer element of histone pre-mRNA (Bond et al. 1991). Modified U7 snRNPs containing an antisense sequence against a β-globin cryptic splice site, several DMD exons and cyclophilin A, Tat or Rev (proteins involved in HIV multiplication) modulated splicing of targeted genes (Suter et al. 1999; Goyenvalle et al. 2004; Liu et al. 2004; Asparuhova et al. 2007). Alternatively, using bifunctional U7 snRNAs containing both an antisense and a splicing enhancer sequence inclusion of *SMN2* exon 7 could be established (Marquis et al. 2007). As the *U7* gene is small, it fits easily in adeno-associated virus (AAV) vectors, which is one of the few viral vectors that can efficiently infect muscle cells (Blankinship et al. 2004). Intramuscular and systemic treatment with AAV vectors containing U7 antisense constructs have shown promising results in the *mdx* mouse and golden retriever muscular dystrophy models (Goyenvalle et al. 2004; L. Garcia, pers. comm.). Exon skipping and dystrophin restoration were observed at high levels for at least 18 months and induced functional improvement. An advantage of this approach is that the antisense sequence is expressed for longer periods of time, thus eliminating the need for repeated injections. However, this is at the same time also a disadvantage. Using AONs, the treatment can be stopped in time, e.g., when better target sequences have been developed or when unexpected (long-term) adverse effects are observed. In addition, the use of a viral vector to deliver the antisense sequence converts the genetic exon skipping therapy into gene therapy. This will be accompanied by the typical gene therapy issues such as vector immunity and insertional mutagenesis. Finally, the manufacturing of AONs can be easily scaled up, whereas reproducibly and safely producing high titers of AAV vectors of excellent purity remain a major challenge. Unless these problems are solved, using AONs is probably a better option, especially since, so far, no toxicity has been reported after long-term systemic injections of AONs (Kurreck 2003; Takeshima et al. 2006).

BROADENING THE FIELD OF EXON SKIPPING APPLICATIONS

In addition to applications described in this review, antisense-mediated exon skipping can be implemented in numerous other therapeutic interventions or developmental studies. The number of applications where antisense-induced restoration of the open reading frame will be therapeutic, like for DMD and dystrophic epidermolysis bullosa, is probably limited. Generally, in-frame deletions of one or more exons will not result in partly functional proteins. However, numerous diseases, especially cancer, are associated with changes in the relative levels of alternative splicing (Srebrow and Kornblihtt 2006). Because the exon skipping effect is titratable, AONs can be employed to normalize levels of alternative splicing or to study the effect of disrupting normal splicing pattern. Finally, exon skipping is a useful tool for gene knockdown. In some areas this approach may be preferable over RNAi, which induces a catalytic process resulting in complete gene knockdown. Exon skipping on the other hand can be used to achieve variable, and more subtle and controlled, levels of knockdown, which is advantageous when (near) complete knockdown is detrimental for the cell, or when a certain amount of knockdown is required. Interestingly, this modulation mimics regulated unproductive splicing and translation, a naturally occurring mechanism, where alternative splice forms contain a premature stop codon and through nonsense-mediated RNA decay regulates protein expression levels (Lewis et al. 2003). In addition, AONs offer the opportunity for isoform or allele-specific knockdown. Ironically, antisense-mediated exon skipping is probably applicable to all areas said to benefit from RNase H-mediated knockdown. Thus, in this way AONs finally may be able to fulfill the promises made over a decade ago.

ACKNOWLEDGMENTS

We thank Johan den Dunnen, Judith van Deutekom, and Mark Einderhand for reading the manuscript and helpful comments. A.A.-R. is funded by a grant from ZonMw (The Netherlands).

REFERENCES

Aartsma-Rus, A. and van Deutekom, J.C.T. 2007. Antisense-mediated reading-frame restoration as a genetic therapy for Duchenne muscular dystrophy. In *Antisense elements (genetics)* (ed. A.G. Hernandes). Novapublishers, Hauppauge, NY.

Aartsma-Rus, A., Janson, A.A., Kaman, W.E., Bremmer-Bout, M., den Dunnen, J.T., Baas, F., van Ommen, G.J., and van Deutekom, J.C. 2003. Therapeutic antisense-induced exon skipping in cultured muscle cells from six different DMD patients. *Hum. Mol. Genet.* **12:** 907–914.

Aartsma-Rus, A., Janson, A.A., Kaman, W.E., Bremmer-Bout, M., van Ommen, G.J., den Dunnen, J.T., and van Deutekom, J.C. 2004a. Antisense-induced multiexon skipping for duchenne muscular dystrophy makes more sense. *Am. J. Hum. Genet.* **74:** 83–92.

Aartsma-Rus, A., Kaman, W.E., Bremmer-Bout, M., Janson, A., Den Dunnen, J., van Ommen, G.J., and van Deutekom, J.C. 2004b. Comparative analysis of antisense oligonucleotide analogs for tageted DMD exon 46 skipping in muscle cells. *Gene Ther.* **11:** 1391–1398.

Aartsma-Rus, A., Winter, C.L., Janson, A.A.M., Kaman, W.E., van Ommen, G.J., den Dunnen, J.T., and van Deutekom, J.C. 2005. Functional analysis of 114 exon-internal AONs for targeted DMD

exon skipping: Indication for steric hindrance of SR protein binding sites. *Oligonucleotides* **15:** 284–297.

Aartsma-Rus, A., Janson, A.A., Heemskerk, J.A., De Winter, C.L., van Ommen, G.J., and van Deutekom, J.C. 2006a. Therapeutic modulation of DMD splicing by blocking exonic splicing enhancer sites with antisense oligonucleotides. *Ann. N. Y. Acad. Sci.* **1082:** 74–76.

Aartsma-Rus, A., Kaman, W.E., Weij, R., den Dunnen, J.T., van Ommen, G.J., and van Deutekom, J.C. 2006b. Exploring the frontiers of therapeutic exon skipping for Duchenne muscular dystrophy by double targeting within one or multiple exons. *Mol. Ther.* **14:** 401–407.

Aartsma-Rus, A., van Deutekom, J.C., Fokkema, I.F., van Ommen, G.J., and den Dunnen, J.T. 2006c. Entries in the Leiden Duchenne muscular dystrophy mutation database: An overview of mutation types and paradoxical cases that confirm the reading-frame rule. *Muscle Nerve* **34:** 135–144.

Aartsma-Rus, A., Janson, A.A., van Ommen, G.-J.B., and van Deutekom, J.C.T. 2007. Antisense-induced exon skipping for duplications in Duchenne muscular dystrophy. *BMC Med. Genet.* **8:** 43. doi: 10.1186/1471-2350-8-43.

Alter, J., Lou, F., Rabinowitz, A., Yin, H., Rosenfeld, J., Wilton, S.D., Partridge, T.A., and Lu, Q.L. 2006. Systemic delivery of morpholino oligonucleotide restores dystrophin expression bodywide and improves dystrophic pathology. *Nat. Med.* **12:** 175–177.

Amantana, A. and Iversen, P.L. 2005. Pharmacokinetics and biodistribution of phosphorodiamidate morpholino antisense oligomers. *Curr. Opin. Pharmacol.* **5:** 550–555.

Ars, E., Serra, E., Garcia, J., Kruyer, H., Gaona, A., Lazaro, C., and Estivill, X. 2000. Mutations affecting mRNA splicing are the most common molecular defects in patients with neurofibromatosis type 1. *Hum. Mol. Genet.* **9:** 237–247.

Asparuhova, M.B., Marti, G., Liu, S., Serhan, F., Trono, D., and Schumperli, D. 2007. Inhibition of HIV-1 multiplication by a modified U7 snRNA inducing Tat and Rev exon skipping. *J. Gene Med.* **9:** 323–334.

Beggs, A.H., Koenig, M., Boyce, F.M., and Kunkel, L.M. 1990. Detection of 98% of DMD/BMD gene deletions by polymerase chain reaction. *Hum. Genet.* **86:** 45–48.

Beroud, C., Tuffery-Giraud, S., Matsuo, M., Hamroun, D., Humbertclaude, V., Monnier, N., Moizard, M.P., Voelckel, M.A., Calemard, L.M., Boisseau, P., et al. 2007. Multiexon skipping leading to an artificial DMD protein lacking amino acids from exons 45 through 55 could rescue up to 63% of patients with Duchenne muscular dystrophy. *Hum. Mutat.* **28:** 196–202.

Blankinship, M.J., Gregorevic, P., Allen, J.M., Harper, S.Q., Harper, H., Halbert, C.L., Miller, D.A., and Chamberlain, J.S. 2004. Efficient transduction of skeletal muscle using vectors based on adeno-associated virus serotype 6. *Mol. Ther.* **10:** 671–678.

Bond, U.M., Yario, T.A., and Steitz, J.A. 1991. Multiple processing-defective mutations in a mammalian histone pre-mRNA are suppressed by compensatory changes in U7 RNA both in vivo and in vitro. *Genes & Dev.* **5:** 1709–1722.

Burns, K., Martinon, F., Esslinger, C., Pahl, H., Schneider, P., Bodmer, J.L., Di Marco, F., French, L., and Tschopp, J. 1998. MyD88, an adapter protein involved in interleukin-1 signaling. *J. Biol. Chem.* **273:** 12203–12209.

Burns, K., Janssens, S., Brissoni, B., Olivos, N., Beyaert, R., and Tschopp, J. 2003. Inhibition of interleukin 1 receptor/Toll-like receptor signaling through the alternatively spliced, short form of MyD88 is due to its failure to recruit IRAK-4. *J. Exp. Med.* **197:** 263–268.

Cartegni, L. and Krainer, A.R. 2002. Disruption of an SF2/ASF-dependent exonic splicing enhancer in SMN2 causes spinal muscular atrophy in the absence of SMN1. *Nat. Genet.* **30:** 377–384.

Cartegni, L. and Krainer, A.R. 2003. Correction of disease-associated exon skipping by synthetic exon-specific activators. *Nat. Struct. Biol.* **10:** 120–125.

Cartegni, L., Chew, S.L., and Krainer, A.R. 2002. Listening to silence and understanding nonsense: Exonic mutations that affect splicing. *Nat. Rev. Genet.* **3:** 285–298.

Cartegni, L., Wang, J., Zhu, Z., Zhang, M.Q., and Krainer, A.R. 2003. ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic Acids Res.* **31:** 3568–3571. doi: 10.1093/nar/gkg616.

Chester, A., Scott, J., Anant, S., and Navaratnam, N. 2000. RNA editing: Cytidine to uridine conversion in apolipoprotein B mRNA. *Biochim. Biophys. Acta* **1494:** 1–13.

Davis, M.I., Bennett, M.J., Thomas, L.M., and Bjorkman, P.J. 2005. Crystal structure of prostate-specific membrane antigen, a tumor marker and peptidase. *Proc. Natl. Acad. Sci.* **102:** 5981–5986.

Deburgrave, N., Daoud, F., Llense, S., Barbot, J.C., Récan, D., Peccate, C., Burghes, A.H., Béroud, C., Garcia, L., Kaplan, J.C., et al. 2007. Protein- and mRNA-based phenotype-genotype correlations in DMD/BMD with point mutations and molecular basis for BMD with nonsense and frameshift mutations in the DMD gene. *Hum. Mutat.* **28:** 183–195.

Disset, A., Bourgeois, C.F., Benmalek, N., Claustres, M., Stevenin, J., and Tuffery-Giraud, S. 2006. An exon skipping-associated nonsense mutation in the dystrophin gene uncovers a complex interplay between multiple antagonistic splicing elements. *Hum. Mol. Genet.* **15:** 999–1013.

Dominski, Z. and Kole, R. 1993. Restoration of correct splicing in thalassemic pre-mRNA by antisense oligonucleotides. *Proc. Natl. Acad. Sci.* **90:** 8673–8677.

Dunckley, M.G., Manoharan, M., Villiet, P., Eperon, I.C., and Dickson, G. 1998. Modification of splicing in the dystrophin gene in cultured Mdx muscle cells by antisense oligoribonucleotides. *Hum. Mol. Genet.* **7:** 1083–1090.

Elbashir, S.M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., and Tuschl, T. 2001. Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411:** 494–498.

Emery, A.E. 2002. The muscular dystrophies. *Lancet* **359:** 687–695.

England, S.B., Nicholson, L.V., Johnson, M.A., Forrest, S.M., Love, D.R., Zubrzycka-Gaarn, E.E., Bulman, D.E., Harris, J.B., and Davies, K.E. 1990. Very mild muscular dystrophy associated with the deletion of 46% of dystrophin. *Nature* **343:** 180–182.

Fairbrother, W.G., Yeh, R.F., Sharp, P.A., and Burge, C.B. 2002. Predictive identification of exonic splicing enhancers in human genes. *Science* **297:** 1007–1013.

Fairbrother, W.G., Holste, D., Burge, C.B., and Sharp, P.A. 2004. Single nucleotide polymorphism-based validation of exonic splicing enhancers. *PLoS Biol.* **2:** e268. doi: 10.1371/journal.pbio.0020268.

Fall, A.M., Johnsen, R., Honeyman, K., Iversen, P., Fletcher, S., and Wilton, S.D. 2006. Induction of revertant fibres in the mdx mouse using antisense oligonucleotides. *Genet. Vaccines Ther.* **4:** 3. doi: 10.1186/1479-0556-4-3.

Fine, J.D., Eady, R.A., Bauer, E.A., Briggaman, R.A., Bruckner-Tuderman, L., Christiano, A., Heagerty, A., Hintner, H., Jonkman, M.F., McGrath, J., et al. 2000. Revised classification system for inherited epidermolysis bullosa: Report of the Second International Consensus Meeting on diagnosis and classification of epidermolysis bullosa. *J. Am. Acad. Dermatol.* **42:** 1051–1066.

Fletcher, S., Honeyman, K., Fall, A.M., Harding, P.L., Johnsen, R.D., and Wilton, S.D. 2006. Dystrophin expression in the *mdx* mouse after localised and systemic administration of a morpholino antisense oligonucleotide. *J. Gene Med.* **8:** 207–216.

Fletcher, S., Honeyman, K., Fall, A.M., Harding, P.L., Johnsen, R.D., Steinhaus, J.P., Moulton, H.M., Iversen, P.L., and Wilton, S.D. 2007. Morpholino oligomer-mediated exon skipping averts the onset of dystrophic pathology in the mdx mouse. *Mol. Ther.* doi: 10.1038/sj.mt.6300245.

Fluiter, K., ten Asbroek, A.L., de Wissel, M.B., Jakobs, M.E., Wissenbach, M., Olsson, H., Olsen, O., Oerum, H., and Baas, F. 2003. In vivo tumor growth inhibition and biodistribution studies

of locked nucleic acid (LNA) antisense oligonucleotides. *Nucl. Acids Res.* **31:** 953–962. doi: 10.1093/nar/gkg185.

Freiburg, A., Trombitas, K., Hell, W., Cazorla, O., Fougerousse, F., Centner, T., Kolmerer, B., Witt, C., Beckmann, J.S., Gregorio, C.C., et al. 2000. Series of exon-skipping events in the elastic spring region of titin as the structural basis for myofibrillar elastic diversity. *Circ. Res.* **86:** 1114–1121.

Friedman, K.J., Kole, J., Cohn, J.A., Knowles, M.R., Silverman, L.M., and Kole, R. 1999. Correction of aberrant splicing of the cystic fibrosis transmembrane conductance regulator (CFTR) gene by antisense oligonucleotides. *J. Biol. Chem.* **274:** 36193–36199.

Gebski, B.L., Mann, C.J., Fletcher, S., and Wilton, S.D. 2003. Morpholino antisense oligonucleotide induced dystrophin exon 23 skipping in mdx mouse muscle. *Hum. Mol. Genet.* **12:** 1801–1811.

Ghosh, D. and Barry, M.A. 2005. Selection of muscle-binding peptides from context-specific peptide-presenting phage libraries for adenoviral vector targeting. *J. Virol.* **79:** 13667–13672.

Ginjaar, I.B., Kneppers, A.L., v d Meulen, J.D., Anderson, L.V., Bremmer-Bout, M., van Deutekom, J.C., Weegenaar, J., den Dunnen, J.T., and Bakker, E. 2000. Dystrophin nonsense mutation induces different levels of exon 29 skipping and leads to variable phenotypes within one BMD family. *Eur. J. Hum. Genet.* **8:** 793–796.

Goedert, M., Spillantini, M.G., Jakes, R., Rutherford, D., and Crowther, R.A. 1989. Multiple isoforms of human microtubule-associated protein tau: Sequences and localization in neurofibrillary tangles of Alzheimer's disease. *Neuron* **3:** 519–526.

Goto, M., Sawamura, D., Nishie, W., Sakai, K., McMillan, J.R., Akiyama, M., and Shimizu, H. 2006. Targeted skipping of a single exon harboring a premature termination codon mutation: Implications and potential for gene correction therapy for selective dystrophic epidermolysis bullosa patients. *J. Invest. Dermatol.* **126:** 2614–2620.

Goyenvalle, A., Vulin, A., Fougerousse, F., Leturcq, F., Kaplan, J.C., Garcia, L., and Danos, O. 2004. Rescue of dystrophic muscle through U7 snRNA-mediated exon skipping. *Science* **306:** 1796–1799.

Harding, P.L., Fall, A.M., Honeyman, K., Fletcher, S., and Wilton, S.D. 2007. The influence of antisense oligonucleotide length on dystrophin exon skipping. *Mol. Ther.* **15:** 157–166.

Hausen, P. and Stein, H. 1970. Ribonuclease H—An enzyme degrading the RNA moiety of DNA-RNA hybrids. *Eur. J. Biochem.* **14:** 278–283.

Hoffman, E.P., Brown Jr., R.H., and Kunkel, L.M. 1987. Dystrophin: The protein product of the Duchenne muscular dystrophy locus. *Cell* **51:** 919–928.

Hoffman, E.P., Fischbeck, K.H., Brown, R.H., Johnson, M., Medori, R., Loike, J.D., Harris, J.B., Waterston, R., Brooke, M., Specht, L., et al. 1988. Characterization of dystrophin in muscle-biopsy specimens from patients with Duchenne's or Becker's muscular dystrophy. *N. Engl. J. Med.* **318:** 1363–1368.

Hua, Y., Vickers, T.A., Baker, B.F., Bennett, C.F., and Krainer, A.R. 2007. Enhancement of SMN2 exon 7 inclusion by antisense oligonucleotides targeting the exon. *PLoS Biol.* **5:** e73. doi: 10.1371/journal.pbio.0050073.

Jablonka, S., Rossoll, W., Schrank, B., and Sendtner, M. 2000. The role of SMN in spinal muscular atrophy. *J. Neurol.* **247:** 37–42.

Kalbfuss, B., Mabon, S.A., and Misteli, T. 2001. Correction of alternative splicing of tau in frontotemporal dementia and parkinsonism linked to chromosome 17. *J. Biol. Chem.* **276:** 42986–42993.

Karras, J.G., McKay, R.A., Dean, N.M., and Monia, B.P. 2000. Deletion of individual exons and induction of soluble murine interleukin-5 receptor-α chain expression through antisense oligonucleotide-mediated redirection of pre-mRNA splicing. *Mol. Pharmacol.* **58:** 380–387.

Karras, J.G., Maier, M.A., Lu, T., Watt, A., and Manoharan, M. 2001. Peptide nucleic acids are potent modulators of endogenous pre-mRNA splicing of the murine interleukin-5 receptor-alpha chain. *Biochemistry* **40:** 7853–7859.

Khoo, B., Roca, X., Chew, S.L., and Krainer, A.R. 2007. Antisense oligonucleotide-induced alternative splicing of the *APOB* mRNA generates a novel isoform of APOB. *BMC Mol. Biol.* **8:** 3. doi: 10.1186/1471-2199-8-3.

Koenig, M., Monaco, A.P., and Kunkel, L.M. 1988. The complete sequence of dystrophin predicts a rod-shaped cytoskeletal protein. *Cell* **53:** 219–226.

Kole, R. and Sazani, P. 2001. Antisense effects in the cell nucleus: Modification of splicing. *Curr. Opin. Mol. Ther.* **3:** 229–234.

Kolonin, M.G., Sun, J., Do, K.A., Vidal, C.I., Ji, Y., Baggerly, K.A., Pasqualini, R., and Arap, W. 2006. Synchronous selection of homing peptides for multiple tissues by in vivo phage display. *FASEB J.* **20:** 979–981.

Krawczak, M., Reiss, J., and Cooper, D.N. 1992. The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: Causes and consequences. *Hum. Genet.* **90:** 41–54.

Kurreck, J. 2003. Antisense technologies. Improvement through novel chemical modifications. *Eur. J. Biochem.* **270:** 1628–1644.

Lewis, B.P., Green, R.E., and Brenner, S.E. 2003. Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proc. Natl. Acad. Sci.* **100:** 189–192.

Liechti-Gallati, S., Koenig, M., Kunkel, L.M., Frey, D., Boltshauser, E., Schneider, V., Braga, S., and Moser, H. 1989. Molecular deletion patterns in Duchenne and Becker type muscular dystrophy. *Hum. Genet.* **81:** 343–348.

Liu, S., Asparuhova, M., Brondani, V., Ziekau, I., Klimkait, T., and Schumperli, D. 2004. Inhibition of HIV-1 multiplication by antisense U7 snRNAs and siRNAs targeting cyclophilin A. *Nucleic Acids Res.* **32:** 3752–3759. doi: 10.1093/nar/gkh715.

Lu, Q.L., Mann, C.J., Lou, F., Bou-Gharios, G., Morris, G.E., Xue, S.A., Fletcher, S., Partridge, T.A., and Wilton, S.D. 2003. Functional amounts of dystrophin produced by skipping the mutated exon in the *mdx* dystrophic mouse. *Nat. Med.* **8:** 1009–1014.

Mann, C.J., Honeyman, K., Cheng, A.J., Ly, T., Lloyd, F., Fletcher, S., Morgan, J.E., Partridge, T.A., and Wilton, S.D. 2001. Antisense-induced exon skipping and synthesis of dystrophin in the *mdx* mouse. *Proc. Natl. Acad. Sci.* **98:** 42–47.

Mann, C.J., Honeyman, K., McClorey, G., Fletcher, S., and Wilton, S.D. 2002. Improved antisense oligonucleotide induced exon skipping in the *mdx* mouse model of muscular dystrophy. *J. Gene Med.* **4:** 644–654.

Marquis, J., Meyer, K., Angehrn, L., Kampfer, S.S., Rothen-Rutishauser, B., and Schumperli, D. 2007. Spinal muscular atrophy: *SMN2* pre-mRNA splicing corrected by a U7 snRNA derivative carrying a splicing enhancer sequence. *Mol. Ther.* doi: 10.1038/sj.mt.6300200.

Marwick, C. 1998. First "antisense" drug will treat CMV retinitis. *JAMA* **280:** 871.

Matsumura, K. and Campbell, K.P. 1994. Dystrophin-glycoprotein complex: Its role in the molecular pathogenesis of muscular dystrophies. *Muscle Nerve* **17:** 2–15.

Matsuo, M., Masumura, T., Nakajima, T., Kitoh, Y., Takumi, T., Nishio, H., Koga, J., and Nakamura, H. 1990. A very small frame-shifting deletion within exon 19 of the Duchenne muscular dystrophy gene. *Biochem. Biophys. Res. Commun.* **170:** 963–967.

Matsuo, M., Masumura, T., Nishio, H., Nakajima, T., Kitoh, Y., Takumi, T., Koga, J., and Nakamura, H. 1991. Exon skipping during splicing of dystrophin mRNA precursor due to an intra-exon deletion in the dystrophin gene of Duchenne muscular dystrophy kobe. *J. Clin. Invest.* **87:** 2127–2131.

McClorey, G., Fall, A.M., Moulton, H.M., Iversen, P.L., Rasko, J.E., Ryan, M., Fletcher, S., and Wilton, S.D. 2006a. Induced dystrophin exon skipping in human muscle explants. *Neuromuscul. Disord.* **16:** 583–590.

McClorey, G., Moulton, H.M., Iversen, P.L., Fletcher, S., and Wilton, S.D. 2006b. Antisense oligonucleotide-induced exon skipping restores dystrophin expression in vitro in a canine model of DMD. *Gene Ther.* **13:** 1373–1381.

McGrath, J.A., Ashton, G.H., Mellerio, J.E., Salas-Alanis, J.C., Swensson, O., McMillan, J.R., and Eady, R.A. 1999. Moderation of phenotypic severity in dystrophic and junctional forms of epidermolysis bullosa through in-frame skipping of exons containing nonsense or frameshift mutations. *J. Invest. Dermatol.* **113:** 314–321.

Mercatante, D.R., Sazani, P., and Kole, R. 2001. Modification of alternative splicing by antisense oligonucleotides as a potential chemotherapy for cancer and other diseases. *Curr. Cancer Drug Targets* **1:** 211–230.

Mercatante, D.R., Mohler, J.L., and Kole, R. 2002. Cellular response to an antisense-mediated shift of Bcl-x pre-mRNA splicing and antineoplastic agents. *J. Biol. Chem.* **277:** 49374–49382.

Mirabella, M., Galluzzi, G., Manfredi, G., Bertini, E., Ricci, E., De Leo, R., Tonali, P., and Servidei, S. 1998. Giant dystrophin deletion associated with congenital cataract and mild muscular dystrophy. *Neurology* **51:** 592–595.

Monaco, A.P., Bertelson, C.J., Liechti-Gallati, S., Moser, H., and Kunkel, L.M. 1988. An explanation for the phenotypic differences between patients bearing partial deletions of the DMD locus. *Genomics* **2:** 90–95.

Morita, K., Hasegawa, C., Kaneko, M., Tsutsumi, S., Sone, J., Ishikawa, T., Imanishi, T., and Koizumi, M. 2002. 2′-*O*,4′-*C*-ethylene-bridged nucleic acids (ENA): Highly nuclease-resistant and thermodynamically stable oligonucleotides for antisense drug. *Bioorg. Med. Chem. Lett.* **12:** 73–76.

Morita, K., Takagi, M., Hasegawa, C., Kaneko, M., Tsutsumi, S., Sone, J., Ishikawa, T., Imanishi, T., and Koizumi, M. 2003. Synthesis and properties of 2′-*O*,4′-*C*-ethylene-bridged nucleic acids (ENA) as effective antisense oligonucleotides. *Bioorg. Med. Chem.* **11:** 2211–2226.

Munsat, T.L. and Davies, K.E. 1992. International SMA consortium meeting. (26-28 June 1992, Bonn, Germany). *Neuromuscul. Disord.* **2:** 423–428.

Muntoni, F., Bushby, K., and van Ommen, G.-J.B. 2005. 128th ENMC international workshop on "Preclinical optimization and Phase I/II clinical trials using antisense oligonucleotides in Duchenne muscular dystrophy" 22-24 October 2004, Naarden, The Netherlands. *Neuromuscul. Disord.* **15:** 450–457.

O'Byrne, P.M. 2006. Cytokines or their antagonists for the treatment of asthma. *Chest* **130:** 244–250.

O'Keefe, D.S., Bacich, D.J., and Heston, W.D. 2004. Comparative analysis of prostate-specific membrane antigen (PSMA) versus a prostate-specific membrane antigen-like gene. *Prostate* **58:** 200–210.

Obika, S., Nanbu, D., Hari, Y., Andon, J.L., Moria, K.I., Doi, T., and Imanishi, T. 1998. Stability and structural features of the duplexes containing nucleoside analogues with fixed N-type conformation, 2′-*O*,4′-*C*-methyleneribonucleosides. *Tetrahedron Lett.* **39:** 5401–5404.

Panayi, G.S., Corrigall, V.M., and Pitzalis, C. 2001. Pathogenesis of rheumatoid arthritis. The role of T cells and other beasts. *Rheum. Dis. Clin. North Am.* **27:** 317–334.

Pramono, Z.A., Takeshima, Y., Alimsardjono, H., Ishii, A., Takeda, S., and Matsuo, M. 1996. Induction of exon skipping of the dystrophin transcript in lymphoblastoid cells by transfecting an antisense oligodeoxynucleotide complementary to an exon recognition sequence. *Biochem. Biophys. Res. Commun.* **226:** 445–449.

Reddy, J.C. and Licht, J.D. 1996. The WT1 Wilms' tumor suppressor gene: How much do we really know? *Biochim. Biophys. Acta* **1287:** 1–28.

Renshaw, J., Orr, R.M., Walton, M.I., Te, P.R., Williams, R.D., Wancewicz, E.V., Monia, B.P., Workman, P., and Pritchard-Jones, K. 2004. Disruption of WT1 gene expression and exon 5 splicing following cytotoxic drug treatment: Antisense down-regulation of exon 5 alters target gene expression and inhibits cell survival. *Mol. Cancer Ther.* **3:** 1467–1484.

Robberson, B.L., Cote, G.J., and Berget, S.M. 1990. Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol. Cell. Biol.* **10:** 84–94.

Roberts, R.G., Coffey, A.J., Bobrow, M., and Bentley, D.R. 1993. Exon structure of the human dystrophin gene. *Genomics* **16:** 536–538.

Roberts, R.G., Gardner, R.J., and Bobrow, M. 1994. Searching for the 1 in 2,400,000: A review of dystrophin gene point mutations. *Hum. Mutat.* **4:** 1–11.

Roberts, J., Palma, E., Sazani, P., Orum, H., Cho, M., and Kole, R. 2006. Efficient and persistent splice switching by systemically delivered LNA oligonucleotides in mice. *Mol. Ther.* **14:** 471–475.

Samoylova, T.I. and Smith, B.F. 1999. Elucidation of muscle-binding peptides by phage display screening. *Muscle Nerve* **22:** 460–466.

Sazani, P., Gemignani, F., Kang, S.H., Maier, M.A., Manoharan, M., Persmark, M., Bortner, D., and Kole, R. 2002. Systemically delivered antisense oligomers upregulate gene expression in mouse tissues. *Nat. Biotechnol.* **20:** 1228–1233.

Scaffidi, P. and Misteli, T. 2005. Reversal of the cellular phenotype in the premature aging disease Hutchinson–Gilford progeria syndrome. *Nat. Med.* **11:** 440–445.

Scharnhorst, V., van der Eb, A.J., and Jochemsen, A.G. 2001. WT1 proteins: Functions in growth and differentiation. *Gene* **273:** 141–161.

Schwarze, U., Starman, B.J., and Byers, P.H. 1999. Redefinition of exon 7 in the COL1A1 gene of type I collagen by an intron 8 splice-donor-site mutation in a form of osteogenesis imperfecta: Influence of intron splice order on outcome of splice-site mutation. *Am. J. Hum. Genet.* **65:** 336–344.

Seeley, M., Huang, W., Chen, Z., Wolff, W.O., Lin, X., and Xu, X. 2007. Depletion of zebrafish titin reduces cardiac contractility by disrupting the assembly of Z-discs and A-bands. *Circ. Res.* **100:** 238–245.

Shiga, N., Takeshima, Y., Sakamoto, H., Inoue, K., Yokota, Y., Yokoyama, M., and Matsuo, M. 1997. Disruption of the splicing enhancer sequence within exon 27 of the dystrophin gene by a nonsense mutation induces partial skipping of the exon and is responsible for Becker muscular dystrophy. *J. Clin. Invest.* **100:** 2204–2210.

Sicinski, P., Geng, Y., Ryder-Cook, A.S., Barnard, E.A., Darlison, M.G., and Barnard, P.J. 1989. The molecular basis of muscular dystrophy in the *mdx* mouse: A point mutation. *Science* **244:** 1578–1580.

Sierakowska, H., Sambade, M.J., Agrawal, S., and Kole, R. 1996. Repair of thalassemic human β-globin mRNA in mammalian cells by antisense oligonucleotides. *Proc. Natl. Acad. Sci.* **93:** 12840–12844.

Skordis, L.A., Dunckley, M.G., Yue, B., Eperon, I.C., and Muntoni, F. 2003. Bifunctional antisense oligonucleotides provide a *trans*-acting splicing enhancer that stimulates SMN2 gene expression in patient fibroblasts. *Proc. Natl. Acad. Sci.* **100:** 4114–4119.

Smith, P.J., Zhang, C., Wang, J., Chew, S.L., Zhang, M.Q., and Krainer, A.R. 2006. An increased specificity score matrix for the prediction of SF2/ASF-specific exonic splicing enhancers. *Hum. Mol. Genet.* **15:** 2490–2508.

Soutar, A.K. and Naoumova, R.P. 2007. Mechanisms of disease: Genetic causes of familial hypercholesterolemia. *Nat. Clin. Pract. Cardiovasc. Med.* **4:** 214–225.

Spillantini, M.G. and Goedert, M. 1998. Tau protein pathology in neurodegenerative diseases. *Trends Neurosci.* **21:** 428–433.

Srebrow, A. and Kornblihtt, A.R. 2006. The connection between splicing and cancer. *J. Cell Sci.* **119:** 2635–2641.

Stojdl, D.F. and Bell, J.C. 1999. SR protein kinases: The splice of life. *Biochem. Cell Biol.* **77:** 293–298.

Su, S.L., Huang, I.P., Fair, W.R., Powell, C.T., and Heston, W.D. 1995. Alternatively spliced variants of prostate-specific membrane antigen RNA: Ratio of expression as a potential measurement of progression. *Cancer Res.* **55:** 1441–1443.

Summerton, J. and Weller, D. 1997. Morpholino antisense oligomers: Design, preparation, and properties. *Antisense Nucleic Acid Drug Dev.* **7:** 187–195.

Surono, A., Van Khanh, T., Takeshima, Y., Wada, H., Yagi, M., Takagi, M., Koizumi, M., and Matsuo, M. 2004. Chimeric RNA/ethylene-bridged nucleic acids promote dystrophin expression in myocytes of duchenne muscular dystrophy by inducing skipping of the nonsense mutation-encoding exon. *Hum. Gene Ther.* **15:** 749–757.

Suter, D., Tomasini, R., Reber, U., Gorman, L., Kole, R., and Schumperli, D. 1999. Double-target antisense U7 snRNAs promote efficient skipping of an aberrant exon in three human β-thalassemic mutations. *Hum. Mol. Genet.* **8:** 2415–2423.

Tacke, R. and Manley, J.L. 1995. The human splicing factors ASF/SF2 and SC35 possess distinct, functionally significant RNA binding specificities. *EMBO J.* **14:** 3540–3551.

Takeshima, Y., Nishio, H., Sakamoto, H., Nakamura, H., and Matsuo, M. 1995. Modulation of in vitro splicing of the upstream intron by modifying an intra-exon sequence which is deleted from the dystrophin gene in dystrophin Kobe. *J. Clin. Invest.* **95:** 515–520.

Takeshima, Y., Wada, H., Yagi, M., Ishikawa, Y., Minami, R., Nakamura, H., and Matsuo, M. 2001. Oligonucleotides against a splicing enhancer sequence led to dystrophin production in muscle cells from a Duchenne muscular dystrophy patient. *Brain Dev.* **23:** 788–790.

Takeshima, Y., Yagi, M., Wada, H., Ishibashi, K., Nishiyama, A., Kakumoto, M., Sakaeda, T., Saura, R., Okumura, K., and Matsuo, M. 2006. Intravenous infusion of an antisense oligonucleotide results in exon skipping in muscle dystrophin mRNA of Duchenne muscular dystrophy. *Pediatr. Res.* **59:** 690–694.

Tennyson, C.N., Klamut, H.J., and Worton, R.G. 1995. The human dystrophin gene requires 16 hours to be transcribed and is cotranscriptionally spliced. *Nat. Genet.* **9:** 184–190.

Teraoka, S.N., Telatar, M., Becker-Catania, S., Liang, T., Onengut, S., Tolun, A., Chessa, L., Sanal, O., Bernatowska, E., Gatti, R.A., et al. 1999. Splicing defects in the ataxia-telangiectasia gene, ATM: Underlying mutations and consequences. *Am. J. Hum. Genet.* **64:** 1617–1631.

Tuffery-Giraud, S., Saquet, C., Chambert, S., Echenne, B., Marie, C.J., Rivier, F., Cossee, M., Philippe, C., Monnier, N., Bieth, E., et al. 2004. The role of muscle biopsy in analysis of the dystrophin gene in Duchenne muscular dystrophy: Experience of a national referral center. *Neuromuscul. Disord.* **14:** 650–658.

Uitto, J., Hovnanian, A., and Christiano, A.M. 1995. Premature termination codon mutations in the type VII collagen gene (COL7A1) underlie severe recessive dystrophic epidermolysis bullosa. *Proc. Assoc. Am. Physicians* **107:** 245–252.

van Deutekom, J.C. and van Ommen, G.J. 2003. Advances in Duchenne muscular dystrophy gene therapy. *Nat. Rev. Genet.* **4:** 774–783.

van Deutekom, J.C., Bremmer-Bout, M., Janson, A.A., Ginjaar, I.B., Baas, F., den Dunnen, J.T., and van Ommen, G.J. 2001. Antisense-induced exon skipping restores dystrophin expression in DMD patient derived muscle cells. *Hum. Mol. Genet.* **10:** 1547–1554.

Varani, L., Hasegawa, M., Spillantini, M.G., Smith, M.J., Murrell, J.R., Ghetti, B., Klug, A., Goedert, M., and Varani, G. 1999. Structure of tau exon 10 splicing regulatory element RNA and destabilization by mutations of frontotemporal dementia and parkinsonism linked to chromosome 17. *Proc. Natl. Acad. Sci.* **96:** 8229–8234.

Vickers, T.A., Zhang, H., Graham, M.J., Lemonidis, K.M., Zhao, C., and Dean, N.M. 2006. Modification of MyD88 mRNA splicing and inhibition of IL-1β signaling in cell culture and in mice with a 2′-*O*-methoxyethyl-modified oligonucleotide. *J. Immunol.* **176:** 3652–3661.

Wahlestedt, C., Salmi, P., Good, L., Kela, J., Johnsson, T., Hokfelt, T., Broberger, C., Porreca, F., Lai, J., Ren, K., et al. 2000. Potent and nontoxic antisense oligonucleotides containing locked nucleic acids. *Proc. Natl. Acad. Sci.* **97:** 5633–5638.

White, S.J., Aartsma-Rus, A., Flanigan, K.M., Weiss, R.B., Kneppers, A.L., Lalic, T., Janson, A.A., Ginjaar, H.B., Breuning, M.H., and den Dunnen, J.T. 2006. Duplications in the DMD gene. *Hum. Mutat.* **27:** 938–945.

Whitfield, A.J., Barrett, P.H., van Bockxmeer, F.M., and Burnett, J.R. 2004. Lipid disorders and mutations in the *APOB* gene. *Clin. Chem.* **50:** 1725–1732.

Williams, T. and Kole, R. 2006. Analysis of prostate-specific membrane antigen splice variants in LNCap cells. *Oligonucleotides* **16:** 186–195.

Wilton, S.D. and Fletcher, S. 2005. Antisense oligonucleotides in the treatment of Duchenne muscular dystrophy: Where are we now? *Neuromuscul. Disord.* **15:** 399–402.

Wilton, S.D., Lloyd, F., Carville, K., Fletcher, S., Honeyman, K., Agrawal, S., and Kole, R. 1999. Specific removal of the nonsense mutation from the mdx dystrophin mRNA using antisense oligonucleotides. *Neuromuscul. Disord.* **9:** 330–338.

Wilton, S.D., Fall, A.M., Harding, P.L., McClorey, G., Coleman, C., and Fletcher, S. 2007. Antisense oligonucleotide-induced exon skipping across the human dystrophin gene transcript. *Mol. Ther.* **15:** 1288–1296.

Wimmer, K., Roca, X., Beiglbock, H., Callens, T., Etzler, J., Rao, A.R., Krainer, A.R., Fonatsch, C., and Messiaen, L. 2007. Extensive in silico analysis of NF1 splicing defects uncovers determinants for splicing outcome upon 5′ splice-site disruption. *Hum. Mutat.* **28:** 599–612.

Wu, J.Y. and Maniatis, T. 1993. Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell* **75:** 1061–1070.

Zamecnik, P.C. and Stephenson, M.L. 1978. Inhibition of Rous sarcoma virus replication and cell transformation by a specific oligodeoxynucleotide. *Proc. Natl. Acad. Sci.* **75:** 280–284.

Zhang, X.H. and Chasin, L.A. 2004. Computational definition of sequence motifs governing constitutive exon splicing. *Genes & Dev.* **18:** 1241–1250.

Zuker, M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucl. Acids Res.* **31:** 3406–3415. doi: 10.1093/nar/gkg595.

EXHIBIT 6

Current Pharmaceutical Biotechnology, 2004, *5*, 431-439

Neutrally Charged Phosphorodiamidate Morpholino Antisense Oligomers: Uptake, Efficacy and Pharmacokinetics

Vikram Arora*, Gayathri R. Devi and Patrick L. Iversen

Research and Development, AVI BioPharma, Inc.

Abstract: Antisense technology constitutes development of sequence-specific DNA or RNA analogs that can block the activity of selected single-stranded genetic sequences and offer the potential of high specificity lacking in many current drug treatments. The sequencing of the human genome has greatly increased the potential of this approach. Antisense oligonucleotides, the most commonly used antisense approach, are unmodified or chemically modified single stranded RNA or DNA molecules specifically designed to hybridize to corresponding RNA by Watson-Crick binding. Phosphorodiamidate Morpholino oligomers (PMO) are a novel class of non-ionic antisense agents that inhibit gene expression by binding to RNA and sterically blocking processing or translation. PMOs have shown excellent efficiency and safety profile via various routes of administration in multiple animal and human studies. This review will summarize the preclinical studies with PMOs on the road to their development as therapeutic agents with particular emphasis on *in vivo* biodistribution and pharmacokinetics.

Key Words: Cancer, Genomics, c-myc, Prostate, phosphorothioate, Delivery, Fluorescence, PMO.

INTRODUCTION

Antisense oligonucleotides can be suitably modified in a variety of ways to increase their stability, solubility, specificity and biodistribution [1]. Amongst the early generation of oligonucleotide analogs are the phosphorothioate-modified oligonucleotides (PS-ODNs), which contain a sulphur atom in each internucleotide linkage instead of oxygen, and are nuclease resistant compared to the phosphodiester oligonucleotides [2]. Formivirsen (ISIS Pharmaceuticals Inc.), the first antisense drug to be approved by the United States Food and Drug Administration (FDA) in 1998 is a PS-ODN used for the treatment of cytomegalovirus-induced retinitis in patients with AIDS [3].

Most of the oligonucleotides that have entered clinical trials possess an internucleoside charge [4, 5]. The presence of the ionic character has been observed to be responsible for oligonucleotide binding to various proteins and also provides a reactive center for potential drug interactions [1, 6, 7]. Stimulation of B-lymphocyte proliferation [8], activation of the complement cascade [9], and severe hypotension [10] are some of the detrimental physiological effects seen due to non-specific oligonucleotide and protein interactions. Non-specific interactions with cellular compartments have also obscured the differentiation of antisense effects from those due to oligonucleotide aptamers or chemical interactions between reactive metabolites, metal ions and ionic oligonucleotides. The G-quartet effect is an example in which the antiproliferative effects have been identified to be due to either antisense or aptameric effects of the oligonucleotides [11]. The formation of unique adducts between the reactive metabolite of acetominophen, N-acetyl-4-benzoquinoneimine, and the sulfur in the backbone of PS-ODN has been identified to resemble the phosphotriester frequently associated with anticancer alkylating drugs [12]. The ionic oligonucleotides also have the capacity to efficiently chelate a variety of metal ions including iron, copper and zinc [13] and this could potentially enhance their excretion and complicate the antisense therapeutic action.

PHOSPHORODIAMIDATE MORPHOLINO OLIGOMERS: CHEMISTRY AND MECHANISM OF ACTION

The phosphorodiamidate Morpholino oligomer chemistry (Fig. **1**) seems to overcome some of the shortcomings of the PS-ODN for use in antisense studies owing to their unusual

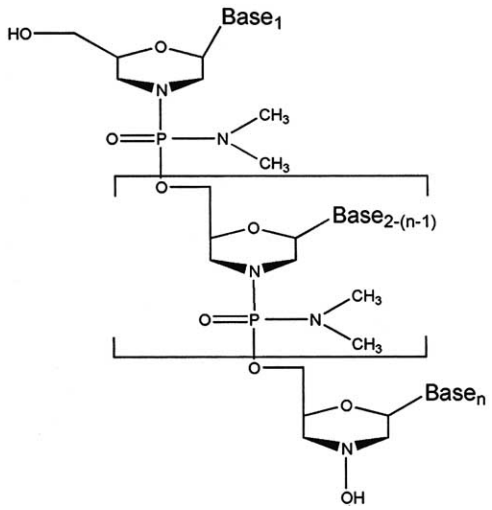


Fig. (1). Structure of phosphorodiamidate Morpholino oligomer.

*Address correspondence to this author at the Research and Development, AVI BioPharma, Inc., 4575 SW Research Way, Corvallis, OR 97333; Tel: (541) 753 3635; Fax: (541) 754 3545; E-mail: varora@avibio.com

1389-2010/04 $45.00+.00

structure, due to replacement of the polynucleotide backbone-5-membered ribose sugars by 6-membered morpholine rings and the phosphodiester intersubunit linkages with phosphorodiamidate linkages [14, 15]. Hence, the PMO are non-ionic and this neutral chemistry avoids a variety of potentially significant limitations observed with the traditional oligonucleotides [16, 17]. The PMOs do not appear to form G-quartets or produce biological effects that are G-quartet related [18]. Unlike the action of PS-ODN analogs, PMO chemistry does not activate RNaseH [19, 20]. Similar to other steric blockers, such as methylphosphonates, peptide nucleic acids, 2'-O-alkyl RNA derivatives, the mechanism of action of PMO involves physically preventing ribosomal assembly thus preventing translation, and interfering with intron-exon splicing of pre-mRNA preventing appropriate translation of selected mRNA [21-23]. The translational arrest mediated by the PMO:mRNA duplex has been evaluated by classical methods in enzymology such as Lineweaver-Burke plots. When the PMO is directed at the 5'UTR the inhibition is non-competitive whereas biphasic, mixed inhibition is observed when directed at the AUG translation initiation start site [24]. The PMOs bind more strongly to complementary RNA than to congenic phosphodiester DNA and show excellent resistance to the action of purified nucleases and proteases found in serum and plasma [25].

DELIVERY AND CELLULAR UPTAKE

The fundamental concern for oligonucleotide-based therapy relies upon the capacity of the oligonucleotide to gain access to the target RNA. The fact that both high molecular weight DNA and RNA have been transferred into mouse muscle cells without cellular uptake enhancers and have remained functional for over 2 months [26] provides a strong historic basis for optimism that synthetic oligomers will enter cells. Indeed, earliest evidence of an oligonucleotide transporter has been described in cultured cells, which is both saturable and promotes cellular uptake [27, 28].

In spite of the fact that there are multiple charged and neutral oligomers in clinical trials, delivery of which is achieved by unassisted systemic administration [2, 4, 5, 29], numerous reports have also demonstrated that naked oligonucleotides are internalized poorly by cells in culture (when delivered unassisted) whether or not they are negatively charged [30-33]. Similarly, although excellent *in vivo* delivery and efficacy of unassisted and non-conjugated PMO has been demonstrated, PMO uptake in cells in culture is mixed and not as robust [34-36]. Several cell culture models are routinely used for rapid primary screening of antisense specificity and activity. PMOs do not efficiently enter most types of cultured cells when added alone, and to achieve a useful level of uptake, it is usually necessary to employ physical or charged-based delivery strategies. Various delivery methods have been optimized based on cell type to assist PMO uptake in culture [34, 37, 38]. Interestingly, recent studies have shown that, unlike well-established cell lines, primary cell cultures seem to internalize PMO without any assisted delivery method similar to the *in vivo* situation [35, 39]. Therefore, identification of contrasts in uptake between cell type and state of differentiation will provide an approach to understanding the molecular mechanisms of cell uptake of oligonucleotides. This will serve to reconcile *in vitro* with *in vivo* observations, provide a basis to compare different oligomer chemistries and lead to optimization of treatment strategies.

The routes of PMO administration have included local, parentral, oral and transdermal, as summarized in Table **1**. In these studies, *in vivo* PMO bioavailability and distribution have been demonstrated either or both qualitatively (fluorescence photomicrography, western immunoblot) and quantitatively (HPLC analysis). A representative in situ photomicrograph is presented in Fig. (**2**) demonstrating intracellular distribution of PMO. PMOs are stable in serum and plasma, but are sensitive to degradation after prolonged exposure to low pH. Previous studies have shown no evidence of truncated versions of PMOs in plasma or liver tissue of Sprague Dawley rats [44].

OLIGONUCLEOTIDE PHARMACOKINETICS

A critical requirement for conducting pharmacokinetic studies is a sensitive and reliable method for quantifying in biological fluids and tissues. Initial observations were primarily with phosphorothioate oligodeoxyribonucleotides that were radiolabeled with ^{35}S [60] which were followed by studies utilizing ^{14}C and ^{3}H modified oligomers. However, "…the apparent biological handling (especially elimination) of oligonucleotides based on radioactivity measurements is influenced by the nature of the radiolabel" [61].

Currently accepted methodologies include capillary gel electrophoresis coupled with mass spectrometry and liquid chromatography coupled with electrospray tandem-mass spectrometry with an ion trap mass spectrometer. Variations of these methods are required for oligomers with different chemical properties such as the charge neutral PMOs which are best evaluated by liquid chromatography followed by detection using a fluorescein-labeled complementary DNA probe.

Pharmacokinetics of Oligonucleotides with Ionic Charge

Pharmacokinetic studies have been completed and reported for phosphodiester, methylphosphonate and a variety of conjugated oligomers (for review see [62]). The most extensively studied pharmacokinetics of any type of synthetic oligonucleotide is the phosphorothioate oligodeoxyribonucleotide (PS-ODN). Initially, PS-ODN were radiolabeled with ^{35}S [60] but as discussed above more sophisticated methods have been developed for detailed pharmacokinetic analysis. A brief review of pharmacokinetic studies of the PS-ODN is summarized in Table **2**.

Studies from different laboratories evaluating the pharmacokinetics of PS-ODNs of different length and with differing methods in a variety of different hosts all tend to observe a similar elimination half-life centered around approximately 30 hours. Evaluation of ISIS 2503 (H-ras), ISIS 5132 (c-raf kinase), ISIS 3521 (PKC-a) and ISIS 2302 (ICAM-1) reveals the same plasma concentration versus time profile during a 2 hour infusion of 1 mg/kg. This observation led scientists at ISIS pharmaceuticals to the conclusion that "…plasma pharmacokinetics and tissue distribution, clearance and ultimately whole-body excretion of PS oligonucleotides have been shown to be sequence indepen-

Table 1. Summary of Literature Demonstrating Phosphorodiamidate Morpholino Activity in Animal Models

Gene Target	Route of Administration	Species	Disease Model	Refe.
Gene Knock-out studies		Developmental models of mouse, Xenopus, chick zebrafish, frog, sea urchin, ascidian	Functional Genomics	[40-43]
c-myc	Intraperitoneal	Rat	Liver regeneration	[44]
p53	Intraperitoneal	Rat	Liver regeneration	[45]
c-myc	Intraperitoneal	Mouse	Polycystic Kidney	[46]
c-myc	Endoluminal catheter	Rabbit, Pig	Restenosis	[47-49]
c-myc	Endoluminal microbubbles	Pig	Restenosis	[50]
c-myc	Intratumoral	Athymic Mouse	Cancer (PC-3 prostate xenograft)	[51]
c-myc	Intraperitoneal	Athymic Mouse	Cancer (DU145 prostate xenograft)	[36]
c-myc	Intraperitoneal	Mouse	Cancer (Syngeneic LLCI tumors)	[52]
β-hCG	Intraperitoneal	Athymic Mouse	Cancer (DU145 prostate xenograft)	[36]
MMP-9	Intratumoral Intraperitoneal	Athymic Mouse	Cancer Angiogenesis	[53]
Androgen Receptor	Intraperitoneal	Athymic Mouse	Cancer	[54], [1]Abstract (Footnote)
α_VIntegrin	Intraperitoneal	Mouse	Retinopathy	[54][1]
TNF-alpha	Pulmonary	Mouse	Pulmonary inflammation	[55]
CYP3A2	Intraperitoneal Intravenous Oral Transdermal	Rat	Drug metabolism	[56-58]
Hepatitis C	Co-delivery with a reporter plasmid	Mice	Hepatitis C	[59]

[1]De Sousa, M.A.; Davies, M.H.; London, C.A.; Powers, M.R.; Devi, G.R. and Farrell, D.H. (2002) 44th American Society of Hematology Proceedings (Abstract).

dent." [71]. Further, these observations can be extended as the plasma clearance rate between 1 and 3 ml/min/kg is largely species independent in rat, rabbit, dog, and monkey. This may be largely due to the observation that more than 96% PS-ODN is bound to plasma proteins in mice, rats, monkey and humans.

The elimination of PS-ODNs is primarily via renal clearance into urine and again this is generally a sequence-independent characteristic. One study indicated fecal excretion was a minor pathway of elimination with 8% excreted in 240 h [69]. In a different presentation, nearly 3.5 % of a 10 mg/kg-administered dose of a PS-ODN was excreted in bile within 4 hours post administration. Further, an oligonucleotide with slower mobility than the administered oligonucleotide could be detected. This product was readily hydrolyzable by glucuronidase, suggesting it is a glucuronic acid conjugate [72].

In summary, the studies with PS-ODNs indicate a relatively long half-life in plasma of about 27-53 hours. They have a wide tissue distribution, mainly in highly perfused tissues, including the liver, kidneys, heart, lungs, and spleen. The brain penetration by PS-ODN is essentially nil. Urinary excretion is the major route of elimination and fecal elimination is less than 10 percent of the administered dose. Oligonucleotides are generally observed as degraded products with smaller molecular weights. Some reports indicate non-linear pharmacokinetics indicating a disproportionate increase in AUC and clearance with dose.

Pharmacokinetics of Phosphorodiamidate Morpholino Oliogmers

There are a number of critical differences between the PMO, which are charge-neutral and the negatively-charged PS-ODN. Pharmacokinetics studies have been conducted in multiple species for three different PMOs to date [73]. A comprehensive evaluation has been completed for AVI-4126, a c-myc antisense PMO which entered human clinical trials early in the year 2000 (Table **3**). The plasma concentration versus time is best fit with a two-compartment model as it is with PS-ODN. The apparent elimination half-life is nearly 10 hours for AVI-4126 (Fig. **3A**). A shorter half-life of 4 and 8 hours is estimated from doses of 0.6 and 1.8 mg/m^2, respectively, but this is believed to be due to the limit of detection by analytical methods. A distribution phase is extended and requires 5 to 10 hours to complete and the

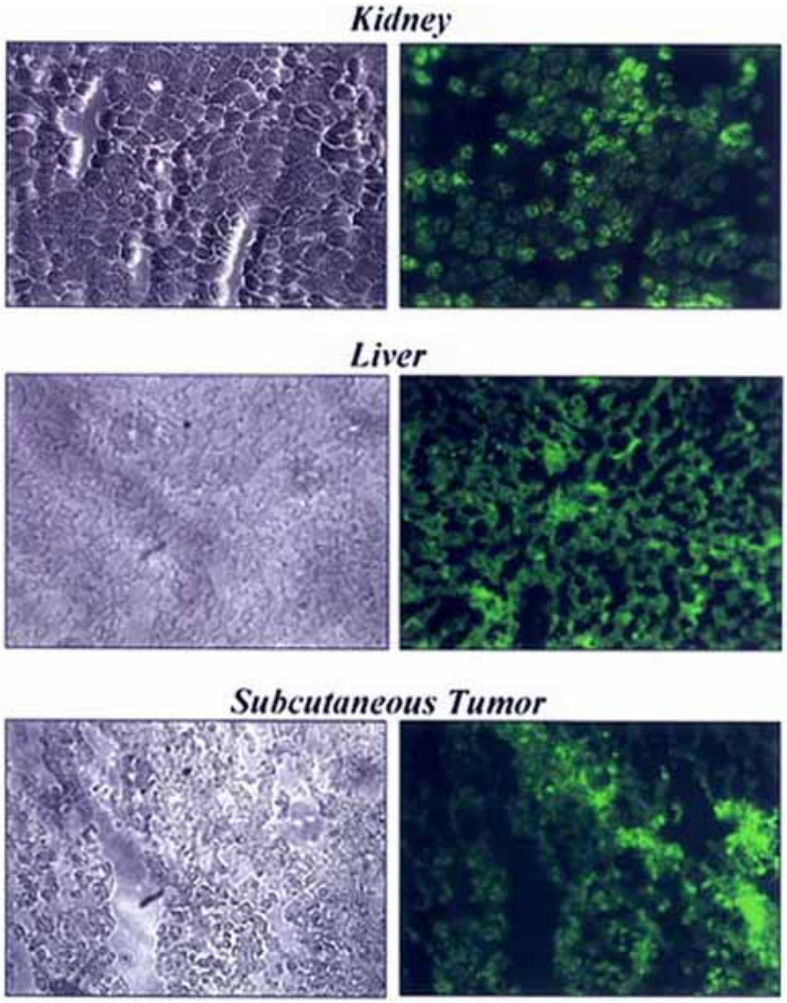


Fig. (2). Representative *in situ* photomicrographs of PMO bioavailability. Phase and fluorescent images of identical fields (200 X) of tissue sections from fluorescein-labeled PMO treated mice (intraperitoneal, 5 mg/kg).

Table 2. Review of Comparable Phosphorothioate Pharmacokinetic Studies

Compound	Dose Mg/kg	Half-life$_{el}$ Hours	Renal Cl % in hours	Reference
HIV-rev	4.9	20-40	>90 in 72 h	[63]
ISIS 2105	3.6	51	15 in 96 h	[64]
ISIS 2105	3.7	53	17 in 240 h	[65]
HIV-rev	0.18-173	27-41	60 in 48 h	[66]
GEM 91	0.1	26.7	49 in 24 h	[67]
GEM 91	30	48	26 in 24 h	[68]
GEM 91	30	48	27 in 24 h 40 in 48 h 58 in 240 h	[69]
OL(1)p53	12-60	25-60	17-59 in 240 h	[70]

Table 3. PMO Agents in Human Clinical Trials

ID	Target	Indication	Phase of Development	Sequence 5'-3'
AVI-4126	*c-myc*	Restenosis	Phase 2 completed	ACGTTG AGG GGC ATC GTC GC
AVI-4126	*c-myc*	Polycyctic Kidney Disease	Phase 1b completed	
AVI-4126	*c-myc*	Solid Tumors	Phase 1 completed	
AVI-4126	*c-myc*	Pharmacokinetics	Phase 1 completed	
AVI-4557	CYP3A4	Drug Metabolism	Phase 1 Completed and in progess	CTG GGA TGA GAG CCA TCA CT
AVI-4020	West Nile	West Nile Viral	Phase 1b in progess	CTTAGACATCGAGATCTTCGT G

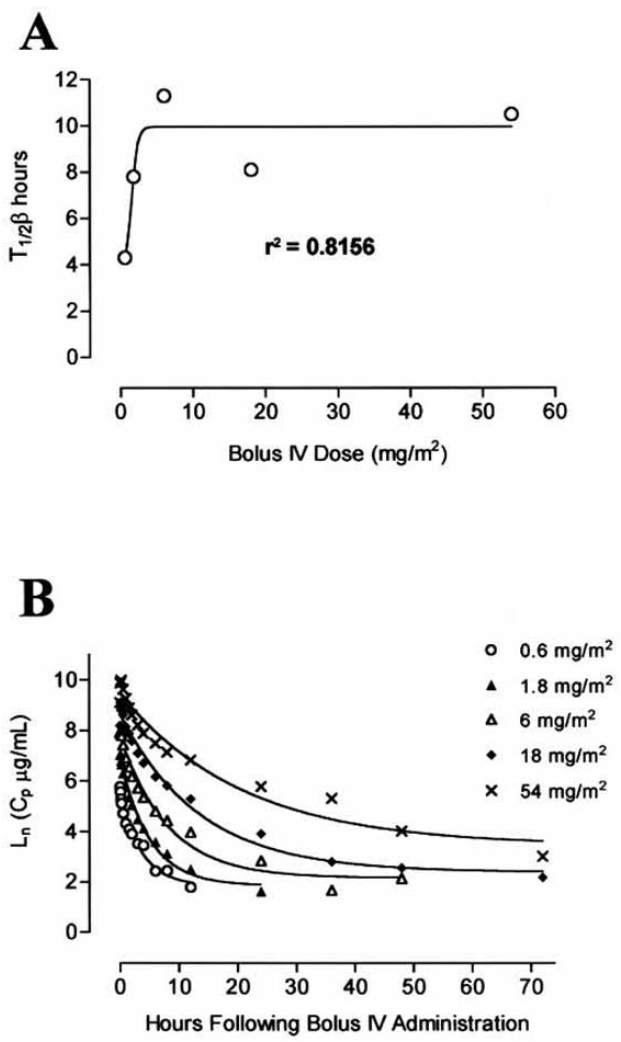


Fig. (3). A) Representative plasma elimination half-life (T1/2 ß) values of AVI-4126 following a single bolus intravenous injection in man. See text for discussion of smaller $T_{1/2}$ ß values at the two lowest doses. B) Representative plasma concentration (Cp) curves of AVI-4126 following a single bolus intravenous injection in man. Note that only full-length oligomer was detected and no degradative metabolism was observed.

number of sample points within the detectable range of the quantitation assay is small for the lower doses (Fig. **3B**). The apparent plasma elimination half-life of AVI-4126 is shorter than the comparable PS-ODN. However, the apparent elimination half-lives of AVI-4557 is longer and AVI-4020 is shorter than AVI-4126. We conclude that unlike the PS-ODNs which have sequence-independent pharmacokinetics, the PMOs demonstrate sequence-specific pharmacokinetic behavior.

The area under the plasma concentration versus time curve (AUC) is linear in man with increasing doses from 0.6 mg/m^2 to 54 mg/m^2 for AVI-4126 (Fig. **4A**). Further, the proportionality of increase in AUC is appropriate for the increase in dose indicating a non-saturable behavior. The plasma clearance for AVI-4126 appears to be constant in humans over nearly 100-fold dose range (Fig. **4B**). The plasma clearance for AVI-4126 is within the reported range for plasma clearance for PS-ODNs. Finally, the AUC is predictable between different animal species including rat, swine and human (Fig. **4C**). We conclude that the pharmacokinetic behavior for a selected PMO is predictable between species and the behavior linear over a clinically relevant dose range. A comparative summary of PS-ODN and PMO are provided in Table **4**.

The distribution of PMOs is primarily to the kidney and liver, which is nearly equivalent to PS-ODNs. However, different PMO sequences influence the degree of tissue accumulation and the relative ratios of organ distribution.

The elimination and distribution of PMOs are also sequence-dependent. The clearance of AVI-4126 into urine has been observed to be between 15 and 28 % of the administered dose within 24 hours of the administered dose in rats and humans. These data are reasonably similar to or slightly smaller than the PS-ODN urinary excretion data. The fecal excretion of the PMO has been determined to be between 10 and 30% in 24 hours and may be sequence-dependent. Less is known about PS-ODN fecal excretion but the published literature indicates less than 10% over a similar period of time. Hence, it is possible that fecal excretion is more equivalent to urinary excretion for PMOs and that fecal excretion is generally greater for PMO than PS-ODN.

PMO EFFICACY STUDIES

Pharmacokinetic measures are not a direct indication of efficacy and judicious interpretation of efficacy should be

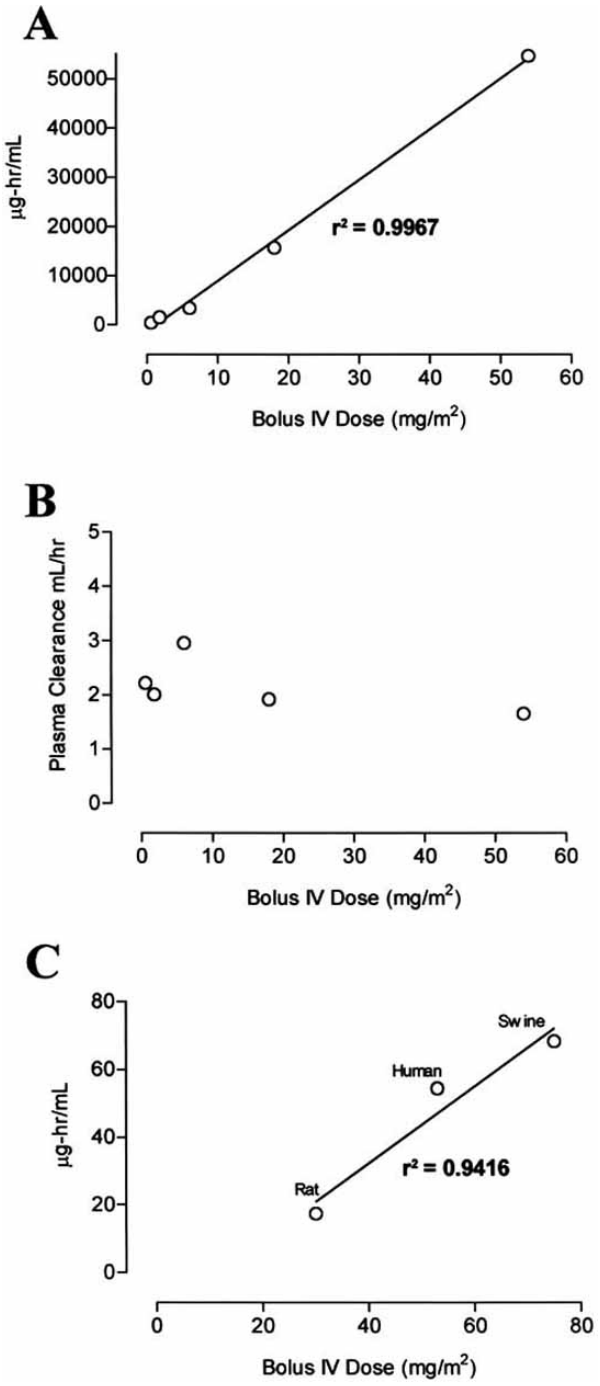


Fig. (4). A) Representative area under plasma concentration curves of AVI-4126 following single bolus intravenous injections in man. B) Representative plasma clearance values of AVI-4126 following single bolus intravenous injections in man. C) Representative area under plasma concentration curves of AVI-4126 following single bolus intravenous injections in rat, swine and man.

linked to pharmacokinetic behavior. PMOs differ from most other existing antisense technologies by combining both neutral charge and non-RNaseH mechanism of action. PMOs have been evaluated in a number of *in vitro* (Table **5**) and *in vivo* models of liver, kidney, lung, cardiovascular, cancer, viral and angiogenesis to confirm antisense efficacy (Table **1**). Further, the observations presented in this review provide insights into the sparse data reconciling pharmacokinetic behavior with *in vivo* efficacy.

Table 4. Pharmacokinetic Comparisons between PS-ODN and PMO

Characteristic	PS-ODN	PMO
Sequence dependence	Sequence-Independent	Sequence-Dependent
Elimination half-life	25-60 hours	<1 to 10 hours
Clearance	1-3 ml/min/kg	1-3 ml/min/kg
Saturable Kinetics	Yes	Sequence-Dependent
Interspecies comparisons	Linear	Linear
Urinary Clearance	26-49% in 24 hours	15-28% in 24 hours
Fecal Elimination	8% in 24 hours	10-30% in 24 hours

The lack of iterative negative charge nearly eliminates non-targeted binding to cellular components other than RNA. Hence, the relative concentration of PMO available for hybridization with RNA is greater. This difference represents the basis for speculation that the sequence-dependent pharmacokinetics for PMOs is unique relative to the iterative ionic oligomers chemistries. Future studies will evaluate the potential that this unique pharmacokinetic behavior reflects relative abundance and constitutive distribution differences in target RNA.

The PMO mechanism of action is also distinct and the temporal onset of action is expected to be immediate based on rapid rate of hybridization with target RNA. In contrast, inhibition that depends upon RNAse activity employs the oligonucleotides as a cofactor in the enzymatic reaction. This substrate-cofactor character may differ for each unique sequence and the k_{cat} rates may be slow relative to the half-life of the target RNA. If the target RNA is relatively high in abundance with relatively short half-life then the enzymatic activity may result in delayed onset of action. We suggest there may be substantial temporal advantages to the direct, non-enzymatic mechanism of action of PMO inhibition of gene expression.

The emerging data presented in this review address two critical aspects pertinent for development of new therapeutic agents. First, *in vitro* methods accurately predict required concentrations for *in vivo* efficacy, as described in Table **6**. Second, while only a few examples are available, the pharmacokinetic behavior observed in animal models accurately predicts human pharmacokinetic behavior. These observations are encouraging in establishing meaningful therapeutic development plans. The critical variables to be isolated with PMO technology include a molecular understanding of the disease to be treated and a more complete understanding of the role of targeted genes in the disease process.

Table 5. Phosphorodiamidate Morpholino Antisense Oligomers Tested *In Vitro*

Target	Disease	Model	Ref.
c-myc	Vascular, Cancer	Primary smooth muscle cells, HeLa, prostate cancer cells	[47, 18, 36, 51]
CYP3A2	Drug Metabolism	Primary hepatocytes	[35]
β-hCG	Cancer	Prostate cancer cells	[36]
MMP-9, MMP-2	Cancer	Prostate cancer cells	[53]
IGF-I Receptor	Cancer	Prostate cancer cells	[54]
XIAP	Cancer	Prostate cancer cells	[1]Abstract (Footnote)
Androgen Receptor	Cancer	Prostate cancer cells	[54], [2]Abstract (Footnote)
Calicivirus	Viral	Viral models of porcine and African monkey kidney cells	[74, 75]
Hepatitic C	Viral	Cell Free System	[76, 77]
TNF-alpha	Cancer	Macrophages, T cells	[78]
Grb2	Vascular Injury	Aortic smooth muscle cells	[79]
β-catenin	Liver development	Embryonic liver cultures	[80]
ROCK−1	Pancreatic Cancer	Pancreatic cancer cell lines	[81]
β-globin	Thalassemia	Erythroid cells	[39, 82]

[1]Amantana, A.; London, C.A.; Iversen, P.L. and Devi, G.R. (2003) Endocrine Society Proceedings (Abstract).
[2]Ko Y.J.; Devi, G.R.; London, C.A.; Kayas, A.; Reddy, M.T.; Iversen, P.L.; Bubley, G. and Balk S.P. (2003) AACR-NCI-EORTC International Conference (Abstract).

Table 6. Correlation of PMO Activity *In Vitro* with *In Vivo* Dose and Efficacy

Target (PMO ID)	In vivo Model	Dose (mg/kg)	Administration Route	% Inhibition *in vivo*	Tissue PMO conc. (24h)	*In vitro* PMO activity (IC_{50} oftarget inhibition)	Ref.
c-myc (AVI-4126)	LLC1 s.c.murine tumor xenograft	5	Intraperitoneal	Abrogation of c-Myc in tumors	57 nM	110 –150 nM	[52]
c-myc (AVI-4126)	PC-3 s.c.murine tumor xenograft	15	Intratumoral	Abrogation of c-Myc in tumors Tumor-static	1280 nM		[51]
c-myc (AVI-4126)	DU145 s.c.murine tumor xenograft	15	Intraperitoneal	Abrogation of c-Myc in tumors	420 nM		[36]
c-myc (AVI-4126)	Porcine Restenosis	0.03	Endoluminal Catheter (Intramural)	>50% c-Myc inhibition in vessels	257 nM		[50]
c-myc (AVI-4126)	Rat Liver Regeneration following 70% partial hepatectomy	2.5	Intraperitoneal	Inhibition of liver regeneration	NA		[44]
PMO-45-1	Mouse prostate	20	Intraperitoneal	Target protein inhibition in prostate	557 nM	626 nM	[1]Abstract in Footnote
PMO-51-3	Glioblastoma s.c. murine xenograft	15	Intraperitoneal	Tumor regression	9 nM	17.3 nM	[2]Abstract in Footnote
AVI-4472 CYP 3A2	Rat Liver	2	Oral	>70 % CYP3A2 enzyme activity in liver	342 nM	275 nM	[56]

[1]Ko Y.J.; Devi, G.R.; London, C.A.; Kayas, A.; Reddy, M.T.; Iversen, P.L.; Bubley, G. and Balk S.P. (2003) AACR-NCI-EORTC International Conference (Abstract).
[2]De Sousa, M.A.; Davies, M.H.; London, C.A.; Powers, M.R.; Devi, G.R. and Farrell, D.H. (2002) 44th American Society of Hematology Proceedings (Abstract).

REFERENCES

[1] Dias, N. and Stein, C.A. (2002) *Mol. Cancer Ther*., **5**, 347-55
[2] Crooke, S.T. (2000) *Methods Enzymol*., **313**, 3-45.
[3] Orr, R.M. (2001) *Curr. Opin. Mol. Ther*., **3**, 288-94.
[4] Devi, G.R. (2002) *Curr. Opin. Mol. Ther*., **4**(2), 138-48.
[5] Pirollo, K.F.; Rait, A.; Sleer, L.S. and Chang, E.H. (2003) *Pharmacol. Ther*., **99**(1), 55-77.
[6] Iversen, P.L.; Copple, B.L. and Tewary, H.K. (1995) *Toxicol. Lett*., **82-83,** 425-30.
[7] Levin, A.A. (1999) *Biochim. Biophys. Acta*., **1489**(1), 69-84.
[8] Kreig, A.M. (1999) *Biochim. Biophys. Acta*., **1489**(1), 107-16
[9] Henry, S.P.; Novotny, W.; Leeds, J.; Auletta, C. and Kornbrust, D.J. (1997) *Antisense Nucleic Acid Drug Dev*., **7**(5), 503-10.
[10] Iversen, P.L.; Cornis, K.G.; Iversen, L.J.; Mata, J.E. and Bylund, D.B. (1999) *Toxicol. Appl. Pharmacol*., **160**(3):289-96.
[11] Burgess, T.L.; Fisher, E.F.; Ross, S.L.; Bready, J.V.; Qian, Y.; Bayewitch, L.A.; Cohen, A.M.; Herrera, C.J.; Hu, S.S.; Kramer, T.B.; Lott, F.D.; Martin, F.H.; Pierce, G.F.; Simonet, L. and Farrell C.L. (1995) *Proc. Natl. Acad. Sci. U S A*, **92**(9), 4051-5.
[12] Copple, B.L.; Gmeiner, W.M. and Iversen, P.L. (1995) *Toxicol. Appl. Pharmacol*., **133**(1), 53-63.
[13] Mata, J.E.; Bishop, M.R.; Tarantolo, S.R.; Angel, C.R.; Swanson, S.A. and Iversen, P.L. (2000) *J. Toxicol. Clin. Toxicol*., **38**(4), 383-7.
[14] Summerton, J. and Weller, D. (1997) *Antisense Nucleic Acid Drug Dev*. **7**(3), 187-95.
[15] Summerton, J. and Weller, D. (1997) *Nucleosides Nucleotides*., **16**(7), 889-898.
[16] Iversen, P.L. (2001) *Curr. Opin. Mol. Ther*., **3**(3), 235-8.
[17] Summerton, J.; Stein, D.; Huang, S.B.; Matthews, P.; Weller, D. and Partridge, M. (1997) *Antisense Nucleic Acid Drug Dev*., **7**(2), 63-70.
[18] Hudziak, R.M.; Summerton, J.; Weller, D.D. and Iversen, P.L. (2000) *Antisense Nucleic Acid Drug Dev*., **10**(3):163-76.
[19] Summerton, J. (1999) *Biochim. Biophys. Acta*., **1489**(1), 141-58.
[20] Stein, C.A. (2000) *Pharmacol. Ther*., **85**(3), 231-6.
[21] Stein, D.; Foster, E.; Huang, S.B.; Weller, D. and Summerton, J. (1997) *Antisense Nucleic Acid Drug Dev*., **7**(3), 151-7.
[22] Giles, R.V.; Spiller, D.G.; Clark, R.E. and Tidd, D.M. (1999) *Antisense Nucleic Acid Drug Dev*., **9**(2), 213-20.
[23] Sazani, P. and Kole, R. (2003*) J. Clin. Invest*., **112**(4), 481-6.
[24] Ghosh, C.; Stein, D.; Weller, D. and Iversen, P.L. *Methods Enzymol*., **313**, 135-43.
[25] Hudziak, R.M.; Barofsky, E.; Barofsky, D.F.; Weller, D.L.; Huang, S.B. and Weller, D.D. (1996) *Antisense Nucleic Acid Drug Dev*., **6**(4), 267-72.
[26] Wolf, J.A.; Malone, R.W. and Williams, P. (1990) *Science*, **247**, 1465-1468.
[27] Feinberg, M.B.; Jarret, R.F.; Aldovini, A; Gallo, R.C. and Wong-Staal, F. (1986) *Cell*, **46,** 807-817.
[28] Sodroski, J.; Goh, W.C.; Rosen, D.; Dayton, A.; Terwilliger, E. and Haseltine, W. (1986) *Nature (London)*., **321**, 412-417.
[29] Stahel, R.A.and Zangemeister-Wittke, U.(2003) *Lung Cancer*., **41 (Suppl 1)**, S81-8.
[30] Stein, C.A. (1999) *Biochim Biophys Acta*., **1489**(1), 45-52.
[31] Tari, A. M. and Lopez-Berestein, G. (2001) *Curr. Opin. Investig. Drugs*., **2**(10), 1450-3.
[32] Thierry, A.R.; Vives, E.; Richard, J.P.; Prevot, P.; Martinand-Mari, C.; Robbins, I. and Lebleu, B. (2003) *Curr. Opin. Mol. Ther*., **5**(2):133-8.
[33] Lebedeva, I.; Benimetskaya, L.; Stein, C.A. and Vilenchik, M. (2000) *Eur. J. Pharm. Biopharm*., **50**(1), 101-19.
[34] Ghosh, C. and Iversen, P.L. (2000) *Antisense Nucleic Acid Drug Dev*., **10**(4), 263-74.
[35] Arora, V.; Cate, M.L.; Ghosh, C. and Iversen, P.L. (2002) *Drug Metab. Dispos*., **30**(7):757-62.
[36] Devi, G.R.; Oldenkamp, J.R.; London, C.A. and Iversen, P.L. (2002) *Prostate*., **53**(3), 200-10.
[37] Partridge, M.; Vincent, A.; Matthews, P.; Puma, J.; Stein, D. and Summerton, J. (1996) *Antisense Nucleic Acid Drug Dev*., **6**(3), 169-75.
[38] Morcos, P.A. (2001) *Genesis*., **30**(3), 94-102.
[39] Suwanmanee, T.; Sierakowska, H.; Lacerra, G.; Svasti, S.; Kirby, S.; Walsh, C.E.; Fucharoen, S. and Kole, R. (2002) *Mol. Pharmacol*., **62**(3), 545-53.
[40] Heasman, J. (2002) *Dev. Biol*., **243**(2), 209-14.
[41] Corey, D.R. and Abrams, J.M. (2001) *Genome Biol*., **2**(5):REVIEWS1015. Epub
[42] Ekker, SC. (2000) *Yeast*., **17**(4), 302-306.
[43] Nasevicius, A. and Ekker, S.C. (2000) *Nat. Genet*., **26**(2), 216-20.
[44] Arora, V.; Knapp, D.C.; Smith, B.L.; Statdfield, M.L.; Stein, D.A.; Reddy, M.T.; Weller, D.D. and Iversen, P.L. (2000) *J. Pharmacol. Exp. Ther*. **292**(3), 921-8.
[45] Arora, V. and Iversen, P.L. (2000) *Drug Metab. Dispos*., **28**(2), 131-8.
[46] Ricker, J.L.; Mata, J.E.; Iversen, P.L. and Gattone, V.H. (2002) *Kidney Int*., **61** (Suppl. 1), 125-131.
[47] Kipshidze, N.; Keane, E.; Stein, D.; Chawla, P.; Skrinska, V.; Shankar, L.R.; Khanna, A.; Komorowski, R.; Haudenschild, C.; Iversen, P.; Leon, M.B.; Keelan, M.H. and Moses, J. (2001) *Catheter Cardiovasc. Interv*., **54**(2), 247-56.
[48] Kipshidze, N.; Moses, J.; Shankar, L.R. and Leon, M. (2001) *Curr. Opin. Mol. Ther*., **3**(3), 265-77.
[49] Kipshidze, N.; Iversen, P.; Keane, E.; Stein, D.; Chawla, P.; Skrinska, V.; Shankar, L.R.; Mehran, R.; Chekanov, V.; Dangas, G.; Komorowski, R.; Haudenschild, C.; Khanna, A.; Leon, M.; Keelan, M.H. and Moses, J. (2002) *Cardiovasc. Radiat. Med*., **3**(1), 26-30.
[50] Kipshidze, N.N.; Kim, H.S.; Iversen, P.; Yazdi, H.A.; Bhargava, B.; New, G.; Mehran, R.; Tio, F.; Haudenschild. C.; Dangas, G.; Stone, G.W.; Iyer, S.; Roubin, G.S.; Leon, M.B. and Moses, J.W. (2002) *J. Am. Coll. Cardiol*., **39**(10), 1686-91.
[51] Iversen, P.L.; Arora, V.; Acker, A.; Mason, D.H. and Devi, G.R. (2003) *Clin. Cancer Res*., 9(7), 2510-9.
[52] Knapp, D.C.; Mata, J.E.; Reddy, M.T.; Devi, G.R. and Iversen, P.L. (2003) *Anticancer Drugs*., **14**(1):39-47.
[53] London, C.A.; Sekhon, H.S.; Arora, V.; Stein, D.A.; Iversen, P.L. and Devi, G.R. (2003) *Cancer Gene Therapy*., in press.
[54] London, C.A.; Knapp, D.C.; Ko, Y.J.; Balk, S.; DeSouza, M.; Farrell, D.H.; Iversen, P.L. and Devi, G.R. (2002) *Cancer Gene Therapy*, **10**(Suppl 1), S3.
[55] Qin, G.; Taylor, M.; Ning, Y.Y.; Iversen, P. and Kobzik, L. (2000) *Antisense Nucleic Acid Drug Dev*., **10**(1), 11-6.
[56] Arora, V.; Knapp, D.C.; Reddy, M.T.; Weller, D.D. and Iversen, P.L. (2002) *J Pharm. Sci*., **91**(4), 1009-18.
[57] Arora, V.; Hannah, T.L.; Iversen, P.L. and Brand, R.M. (2002) *Pharm. Res*., **19**(10), 1465-70.
[58] Arora, V. and Iversen, P.L. (2001) *Curr. Opin. Mol. Ther*. **3**(3), 249-57.
[59] McCaffrey, A.P.; Meuse, L.; Karimi, M.; Contag, C.H. and Kay, M.A. (2003) *Hepatology*, **38**(2), 503-8.
[60] Stein, C.A.; Iversen, P.L.; Subasinghe, C.; Cohen, J.S.; Stec, W.J. and Zon, G. (1990) *Anal. Biochem*., **188,** 11-16
[61] Nicklin, P.L.; Craig, S.J. and Phillips, J.A. (1998) in *Antisense Research and Application* (Crooke, S.T., Ed.), Springer-Verlag, New York, pp. 141-168.
[62] Wallace, T.L.; Cossum, P.A. and Cameron, P.A. (1998) in *Applied Antisense Oligonucleotide Technology*, (Stein, C.A and Krieg, A.M. Eds.) Wiley-Liss, Inc., pp. 407-429.
[63] Iversen, P. (1991) *Anti-Cancer Drug Design*, **6**: 531-538.
[64] Cossum, P.A.; Sasmor, H.; Dellinger, D.; Truong, L.; Cumms, L.; Owens, S.R.; Markham, P.M.; Shea, J.P. and Crooke, S.T. (1993) *J. Pharmacol. Exp. Ther*., **267**, 1181-1190.
[65] Cossum, P.A.; Truong, L.; Owens, S.R.; Markham, P.M.; Shea, J.P. and Crooke, S.T. (1994) *J. Pharmacol Exp. Ther*., **269**, 89-94.
[66] Iversen, P.L.; Mata, J.; Tracewell, W.G. and Zon, G. (1994) *Antisense. Res. Dev.*, **4**, 43-52.
[67] Zhang, R.; Diasio, R.B.; Lu, Z.; Liu, T.; Jiang, Z.; Galbraith, W.M. and Agrawal S. (1995) *Biochem. Pharmacol.*, **49**, 929-939.
[68] Zhang, R.; Lu, Z.; Zhang, X.; Zhao, H.; Diasio, R.B.; Liu, T.; Jiang, Z. and Agrawal, S. (1995) *Clin. Chem*. **41,** 836-843.
[69] Agrawal, S. and Temsamani, J. (1996) in *Methods in Molecular Medicine: Antisense Therapeutics*, (Agrawal, S., Ed.), Humana Press Inc. Totowa, NJ, pp. 247-270.
[70] Bishop, M.R.; Iversen, P.L.; Bayever, E.; Sharp, J.G.; Greiner, T.C.; Copple, B.L.; Ruddon, R.; Zon, G.; Spinolo, J.; Arneson, M.; Armitage, J.O. and Kessinger, A. (1996) *J. Clin. Oncol.*, **14** (4), 1320-6.
[71] Geary, R.S.; Yu, R.Z.; Leeds, J.M.; Watanabe, T.A.; Henry, S.P.; Levin, A.A. and Templin, M.V. (2001) in *Antisense Drug*

Technology Principles, Strategies, and Applications (Crooke, S.T., Ed.), Marcel Dekker, Inc. New York, pp.119-154.

[72] Agrawal, S. and Zhang, R. (1997) *Ciba Foundation Symposium 209*, Wiley, Chichester, pp. 60-78.

[73] Iversen, P.L. (2001) in *Antisense Drug Techcology- Principles, Strategies, and Applications*, (Crooke, S.T., Ed.), Marcel Dekker, Inc., New York, pp. 375-389.

[74] Stein, D.A.; Skilling, D.E.; Iversen, P.L. and Smith, A.W. (2001) *Antisense Nucleic Acid Drug Dev*., **11**(5), 317-25.

[75] Smith, A.W.; Matson, D.O.; Stein, D.A.; Skilling, D.E.; Kroeker, A.D.; Berke, T. and Iversen, P.L. (2002) *Curr. Opin. Mol. Ther*., **4**(2), 177-84.

[76] Jubin, R.; Vantuno, N.E.; Kieft, J.S, ; Murray, M.G.; Doudna, J.A.; Lau, J.Y. and Baroudy, B.M. (2000) *J. Virol*., **74**(22), 10430-7.

[77] Jubin, R. (2001) *Curr. Opin. Mol. Ther*., **3**(3):278-87.

[78] Taylor, M.F.; Weller, D.D. and Kobzik, L. (1998) *Antisense Nucleic Acid Drug Dev*., **8**(3), 199-205.

[79] Zhang, S.; Ren, J.; Khan, M.F.; Cheng, A.M.; Abendschein, D. and Muslin, A.J. (2003) *Arterioscler. Thromb. Vasc. Biol*., **3** [Epub ahead of print].

[80] Monga, S.P.; Monga, H.K.; Tan, X.; Mule, K.; Pediaditakis, P. and Michalopoulos, G.K. (2003) *Gastroenterology*, **124**(1), 202-16.

[81] Kaneko, K.; Satoh, K.; Masamune, A.; Satoh, A. and Shimosegawa, T. (2002) *Pancreas*, **24**(3), 251-7.

[82] Lacerra, G.; Sierakowska, H.; Carestia, C.; Fucharoen, S.; Summerton, J.; Weller, D. and Kole, R. (2000) *Proc. Natl. Acad. Sci. U S A*, **97**(17), 9591-6.

EXHIBIT 7

MOLECULAR BIOLOGY OF

THE CELL

ALBERTS JOHNSON LEWIS RAFF ROBERTS WALTER

FOURTH EDITION

MOLECULAR BIOLOGY OF THE CELL

fourth edition

Bruce Alberts

Alexander Johnson

Julian Lewis

Martin Raff

Keith Roberts

Peter Walter

Garland Science
Taylor & Francis Group

Garland
Vice President: Denise Schanck
Managing Editor: Sarah Gibbs
Senior Editorial Assistant: Kirsten Jenner
Managing Production Editor: Emma Hunt
Proofreader and Layout: Emma Hunt
Production Assistant: Angela Bennett
Text Editors: Marjorie Singer Anderson and Betsy Dilernia
Copy Editor: Bruce Goatly
Word Processors: Fran Dependahl, Misty Landers and Carol Winter
Designer: Blink Studio, London
Illustrator: Nigel Orme
Indexer: Janine Ross and Sherry Granum
Manufacturing: Nigel Eyre and Marion Morrow

Cell Biology Interactive
Artistic and Scientific Direction: Peter Walter
Narrated by: Julie Theriot
Production, Design, and Development: Mike Morales

Bruce Alberts received his Ph.D. from Harvard University and is President of the National Academy of Sciences and Professor of Biochemistry and Biophysics at the University of California, San Francisco. **Alexander Johnson** received his Ph.D. from Harvard University and is a Professor of Microbiology and Immunology and Co-Director of the Biochemistry and Molecular Biology Program at the University of California, San Francisco. **Julian Lewis** received his D.Phil. from the University of Oxford and is a Principal Scientist at the London Research Institute of Cancer Research UK. **Martin Raff** received his M.D. from McGill University and is at the Medical Research Council Laboratory for Molecular Cell Biology and Cell Biology Unit and in the Biology Department at University College London. **Keith Roberts** received his Ph.D. from the University of Cambridge and is Associate Research Director at the John Innes Centre, Norwich. **Peter Walter** received his Ph.D. from The Rockefeller University in New York and is Professor and Chairman of the Department of Biochemistry and Biophysics at the University of California, San Francisco, and an Investigator of the Howard Hughes Medical Institute.

Library of Congress Cataloging-in-Publication Data
Molecular biology of the cell / Bruce Alberts ... [et al.].-- 4th ed.
p. cm
Includes bibliographical references and index.
ISBN 0-8153-3218-1 (hardbound) -- ISBN 0-8153-4072-9 (pbk.)
1. Cytology. 2. Molecular biology. I. Alberts, Bruce.
[DNLM: 1. Cells. 2. Molecular Biology.]
QH581.2 .M64 2002
571.6--dc21

2001054471 CIP

Published by Garland Science, a member of the Taylor & Francis Group, 29 West 35th Street, New York, NY 10001-2299

Printed in the United States of America

15 14 13 12 11 10 9 8 7 6 5 4 3 2

Front cover Human Genome: Reprinted by permission from *Nature,* International Human Genome Sequencing Consortium, 409:860–921, 2001 © Macmillan Magazines Ltd. Adapted from an image by Francis Collins, NHGRI; Jim Kent, UCSC; Ewan Birney, EBI; and Darryl Leja, NHGRI; showing a portion of Chromosome 1 from the initial sequencing of the human genome.

Chapter opener Portion of chromosome 2 from the genome of the fruit fly *Drosophila melanogaster.* (Reprinted with permission from M.D. Adams et al., *Science* 287:2185–2195, 2000. © AAAS.)

Back cover In 1967, the British artist Peter Blake created a design classic. Nearly 35 years later Nigel Orme (illustrator), Richard Denyer (photographer), and the authors have together produced an affectionate tribute to Mr Blake's image. With its gallery of icons and influences, its assembly created almost as much complexity, intrigue and mystery as the original. *Drosophila, Arabidopsis,* Dolly and the assembled company tempt you to dip inside where, as in the original, "a splendid time is guaranteed for all."
(Gunter Blobel, courtesy of The Rockefeller University; Marie Curie, Keystone Press Agency Inc; Darwin bust, by permission of the President and Council of the Royal Society; Rosalind Franklin, courtesy of Cold Spring Harbor Laboratory Archives; Dorothy Hodgkin, © The Nobel Foundation, 1964; James Joyce, etching by Peter Blake; Robert Johnson, photo booth self-portrait early 1930s, © 1986 Delta Haze Corporation all rights reserved, used by permission; Albert L. Lehninger, (unidentified photographer) courtesy of The Alan Mason Chesney Medical Archives of The Johns Hopkins Medical Institutions; Linus Pauling, from Ava Helen and Linus Pauling Papers, Special Collections, Oregon State University; Nicholas Poussin, courtesy of ArtToday.com; Barbara McClintock, © David Micklos, 1983; Andrei Sakharov, courtesy of Elena Bonner; Frederick Sanger, © The Nobel Foundation, 1958.)

Contents

4

DNA AND CHROMOSOMES

Life depends on the ability of cells to store, retrieve, and translate the genetic instructions required to make and maintain a living organism. This *hereditary* information is passed on from a cell to its daughter cells at cell division, and from one generation of an organism to the next through the organism's reproductive cells. These instructions are stored within every living cell as its **genes**, the information-containing elements that determine the characteristics of a species as a whole and of the individuals within it.

As soon as genetics emerged as a science at the beginning of the twentieth century, scientists became intrigued by the chemical structure of genes. The information in genes is copied and transmitted from cell to daughter cell millions of times during the life of a multicellular organism, and it survives the process essentially unchanged. What form of molecule could be capable of such accurate and almost unlimited replication and also be able to direct the development of an organism and the daily life of a cell? What kind of instructions does the genetic information contain? How are these instructions physically organized so that the enormous amount of information required for the development and maintenance of even the simplest organism can be contained within the tiny space of a cell?

The answers to some of these questions began to emerge in the 1940s, when researchers discovered, from studies in simple fungi, that genetic information consists primarily of instructions for making proteins. Proteins are the macromolecules that perform most cellular functions: they serve as building blocks for cellular structures and form the enzymes that catalyze all of the cell's chemical reactions (Chapter 3), they regulate gene expression (Chapter 7), and they enable cells to move (Chapter 16) and to communicate with each other (Chapter 15). The properties and functions of a cell are determined almost entirely by the proteins it is able to make. With hindsight, it is hard to imagine what other type of instructions the genetic information could have contained.

The other crucial advance made in the 1940s was the identification of **deoxyribonucleic acid (DNA)** as the likely carrier of genetic information. But the mechanism whereby the hereditary information is copied for transmission from cell to cell, and how proteins are specified by the instructions in the DNA, remained completely mysterious. Suddenly, in 1953, the mystery was solved when the structure of DNA was determined by James Watson and Francis Crick. As mentioned in Chapter 1, the structure of DNA immediately solved the problem of how the information in this molecule might be copied, or *replicated*. It also provided the first clues as to how a molecule of DNA might encode the instructions for making proteins. Today, the fact that DNA is the genetic material is so fundamental to biological thought that it is difficult to realize what an enormous intellectual gap this discovery filled.

Well before biologists understood the structure of DNA, they had recognized that genes are carried on *chromosomes*, which were discovered in the nineteenth century as threadlike structures in the nucleus of a eucaryotic cell that become visible as the cell begins to divide (Figure 4–1). Later, as biochemical analysis became possible, chromosomes were found to consist of both DNA and protein. We now know that the DNA carries the hereditary information of the cell (Figure 4–2). In contrast, the protein components of chromosomes function largely to package and control the enormously long DNA molecules so that they fit inside cells and can easily be accessed by them.

In this chapter we begin by describing the structure of DNA. We see how, despite its chemical simplicity, the structure and chemical properties of DNA make it ideally suited as the raw material of genes. The genes of every cell on Earth are made of DNA, and insights into the relationship between DNA and genes have come from experiments in a wide variety of organisms. We then consider how genes and other important segments of DNA are arranged on the long molecules of DNA that are present in chromosomes. Finally, we discuss how eucaryotic cells fold these long DNA molecules into compact chromosomes. This packing has to be done in an orderly fashion so that the chromosomes can be replicated and apportioned correctly between the two daughter cells at each cell division. It must also allow access of chromosomal DNA to enzymes that repair it when it is damaged and to the specialized proteins that direct the expression of its many genes.

This is the first of four chapters that deal with basic genetic mechanisms—the ways in which the cell maintains, replicates, expresses, and occasionally improves the genetic information carried in its DNA. In the following chapter (Chapter 5) we discuss the mechanisms by which the cell accurately replicates and repairs DNA; we also describe how DNA sequences can be rearranged through the process of genetic recombination. Gene expression—the process through which the information encoded in DNA is interpreted by the cell to guide the synthesis of proteins—is the main topic of Chapter 6. In Chapter 7, we describe how gene expression is controlled by the cell to ensure that each of the many thousands of proteins encrypted in its DNA is manufactured only at the proper time and place in the life of the cell. Following these four chapters on basic genetic mechanisms, we present an account of the experimental techniques used to study these and other processes that are fundamental to all cells (Chapter 8).

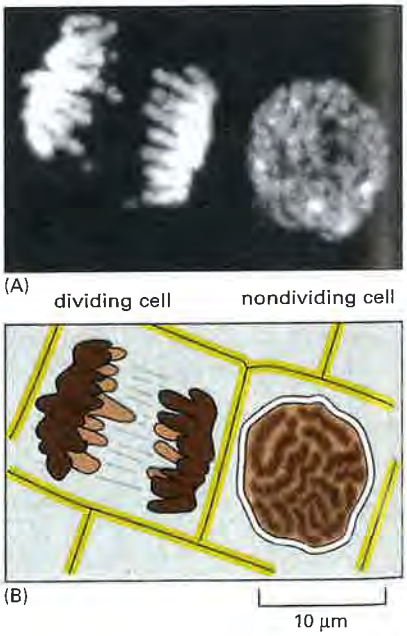


Figure 4–1 Chromosomes in cells. (A) Two adjacent plant cells photographed through a light microscope. The DNA has been stained with a fluorescent dye (DAPI) that binds to it. The DNA is present in chromosomes, which become visible as distinct structures in the light microscope only when they become compact structures in preparation for cell division, as shown on the left. The cell on the right, which is not dividing, contains identical chromosomes, but they cannot be clearly distinguished in the light microscope at this phase in the cell's life cycle, because they are in a more extended conformation. (B) Schematic diagram of the outlines of the two cells along with their chromosomes. (A, courtesy of Peter Shaw.)

THE STRUCTURE AND FUNCTION OF DNA

Biologists in the 1940s had difficulty in accepting DNA as the genetic material because of the apparent simplicity of its chemistry. DNA was known to be a long polymer composed of only four types of subunits, which resemble one another chemically. Early in the 1950s, DNA was first examined by x-ray diffraction analysis, a technique for determining the three-dimensional atomic structure of a molecule (discussed in Chapter 8). The early x-ray diffraction results indicated that DNA was composed of two strands of the polymer wound into a helix. The observation that DNA was double-stranded was of crucial significance and

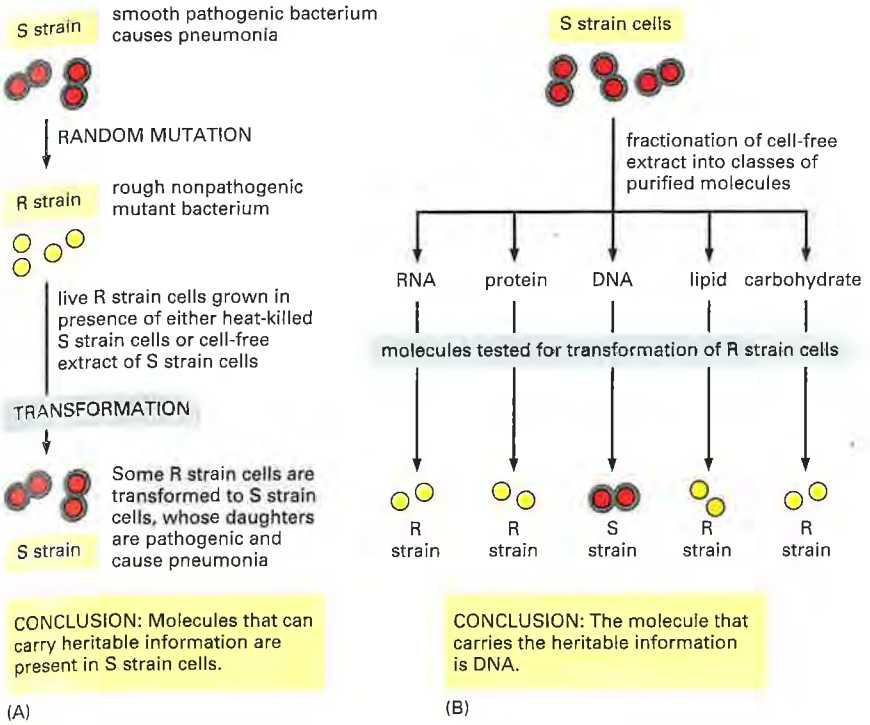


Figure 4–2 Experimental demonstration that DNA is the genetic material. These experiments, carried out in the 1940s, showed that adding purified DNA to a bacterium changed its properties and that this change was faithfully passed on to subsequent generations. Two closely related strains of the bacterium *Streptococcus pneumoniae* differ from each other in both their appearance under the microscope and their pathogenicity. One strain appears smooth (S) and causes death when injected into mice, and the other appears rough (R) and is nonlethal. (A) This experiment shows that a substance present in the S strain can change (or transform) the R strain into the S strain and that this change is inherited by subsequent generations of bacteria. (B) This experiment, in which the R strain has been incubated with various classes of biological molecules obtained from the S strain, identifies the substance as DNA.

provided one of the major clues that led to the Watson–Crick structure of DNA. Only when this model was proposed did DNA's potential for replication and information encoding become apparent. In this section we examine the structure of the DNA molecule and explain in general terms how it is able to store hereditary information.

A DNA Molecule Consists of Two Complementary Chains of Nucleotides

A DNA molecule consists of two long polynucleotide chains composed of four types of nucleotide subunits. Each of these chains is known as a *DNA chain*, or a *DNA strand*. *Hydrogen bonds* between the base portions of the nucleotides hold the two chains together (Figure 4–3). As we saw in Chapter 2 (Panel 2–6, pp. 120-121), nucleotides are composed of a five-carbon sugar to which are attached one or more phosphate groups and a nitrogen-containing base. In the case of the nucleotides in DNA, the sugar is deoxyribose attached to a single phosphate group (hence the name deoxyribonucleic acid), and the base may be either *adenine (A)*, *cytosine (C)*, *guanine (G)*, or *thymine (T)*. The nucleotides are covalently linked together in a chain through the sugars and phosphates, which thus form a "backbone" of alternating sugar–phosphate–sugar–phosphate (see Figure 4–3). Because only the base differs in each of the four types of subunits, each polynucleotide chain in DNA is analogous to a necklace (the backbone) strung with four types of beads (the four bases A, C, G, and T). These same symbols (A, C, G, and T) are also commonly used to denote the four different nucleotides—that is, the bases with their attached sugar and phosphate groups.

The way in which the nucleotide subunits are lined together gives a DNA strand a chemical polarity. If we think of each sugar as a block with a protruding knob (the 5′ phosphate) on one side and a hole (the 3′ hydroxyl) on the other (see Figure 4–3), each completed chain, formed by interlocking knobs with holes, will have all of its subunits lined up in the same orientation. Moreover, the two ends of the chain will be easily distinguishable, as one has a hole (the 3′ hydroxyl) and the other a knob (the 5′ phosphate) at its terminus. This polarity in a DNA chain is indicated by referring to one end as the *3′ end* and the other as the *5′ end*.

The three-dimensional structure of DNA—**the double helix**—arises from the chemical and structural features of its two polynucleotide chains. Because

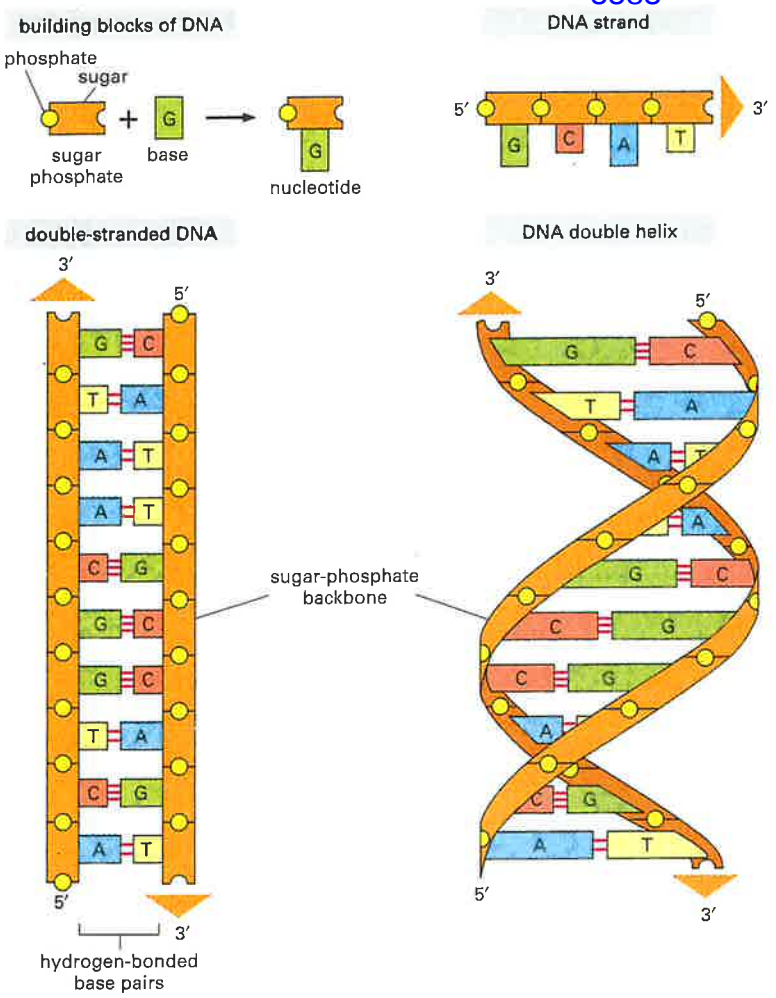


Figure 4–3 DNA and its building blocks. DNA is made of four types of nucleotides, which are linked covalently into a polynucleotide chain (a DNA strand) with a sugar-phosphate backbone from which the bases (A, C, G, and T) extend. A DNA molecule is composed of two DNA strands held together by hydrogen bonds between the paired bases. The *arrowheads* at the ends of the DNA strands indicate the polarities of the two strands, which run antiparallel to each other in the DNA molecule. In the diagram at the bottom left of the figure, the DNA molecule is shown straightened out; in reality, it is twisted into a double helix, as shown on the right. For details, see Figure 4–5.

these two chains are held together by hydrogen bonding between the bases on the different strands, all the bases are on the inside of the double helix, and the sugar-phosphate backbones are on the outside (see Figure 4–3). In each case, a bulkier two-ring base (a purine; see Panel 2–6, pp. 120–121) is paired with a single-ring base (a pyrimidine); A always pairs with T, and G with C (Figure 4–4). This *complementary base-pairing* enables the **base pairs** to be packed in the

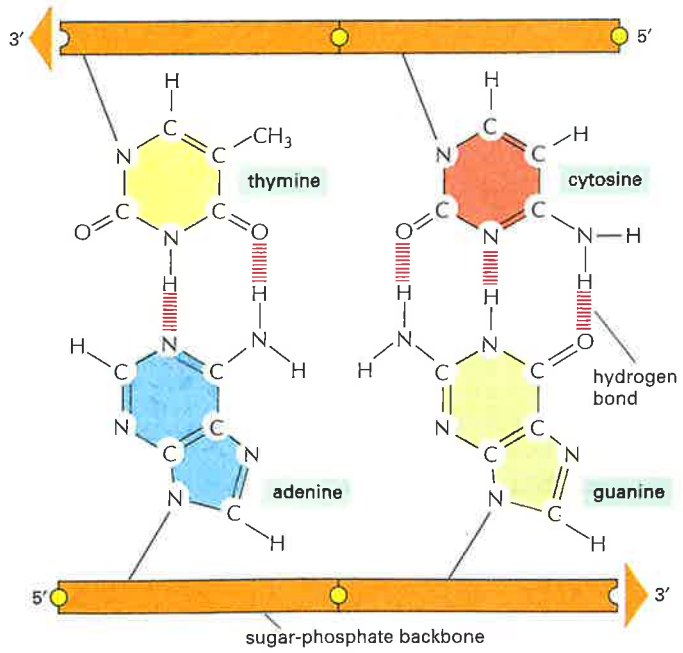


Figure 4–4 Complementary base pairs in the DNA double helix. The shapes and chemical structure of the bases allow hydrogen bonds to form efficiently only between A and T and between G and C, where atoms that are able to form hydrogen bonds (see Panel 2–3, pp. 114–115) can be brought close together without distorting the double helix. As indicated, two hydrogen bonds form between A and T, while three form between G and C. The bases can pair in this way only if the two polynucleotide chains that contain them are antiparallel to each other.

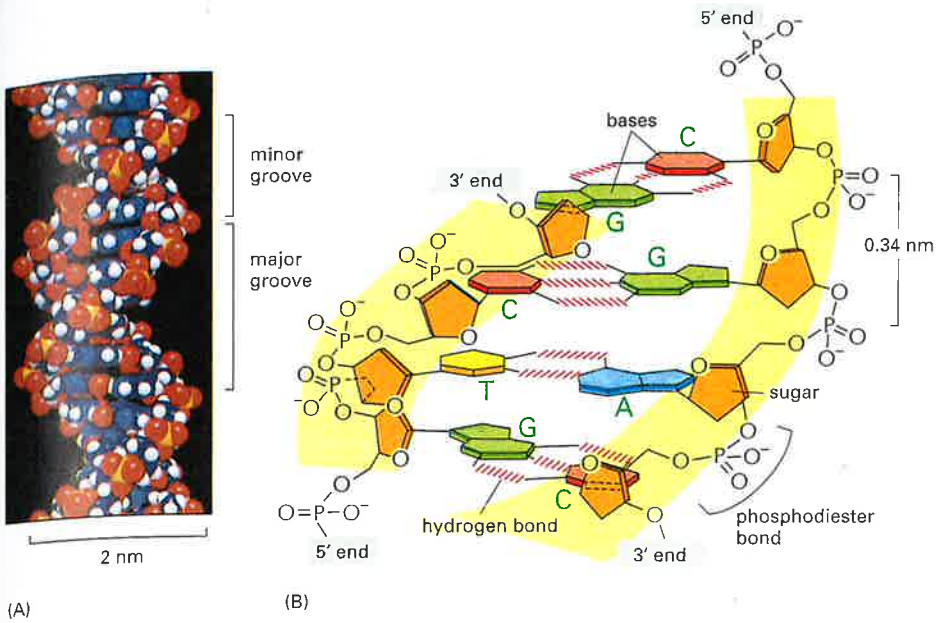


Figure 4–5 The DNA double helix. (A) A space-filling model of 1.5 turns of the DNA double helix. Each turn of DNA is made up of 10.4 nucleotide pairs and the center-to-center distance between adjacent nucleotide pairs is 3.4 nm. The coiling of the two strands around each other creates two grooves in the double helix. As indicated in the figure, the wider groove is called the major groove, and the smaller the minor groove. (B) A short section of the double helix viewed from its side, showing four base pairs. The nucleotides are linked together covalently by phosphodiester bonds through the 3′-hydroxyl (–OH) group of one sugar and the 5′-phosphate (P) of the next. Thus, each polynucleotide strand has a chemical polarity; that is, its two ends are chemically different. The 3′ end carries an unlinked –OH group attached to the 3′ position on the sugar ring; the 5′ end carries a free phosphate group attached to the 5′ position on the sugar ring.

energetically most favorable arrangement in the interior of the double helix. In this arrangement, each base pair is of similar width, thus holding the sugar-phosphate backbones an equal distance apart along the DNA molecule. To maximize the efficiency of base-pair packing, the two sugar-phosphate backbones wind around each other to form a double helix, with one complete turn every ten base pairs (Figure 4–5).

The members of each base pair can fit together within the double helix only if the two strands of the helix are **antiparallel**—that is, only if the polarity of one strand is oriented opposite to that of the other strand (see Figures 4–3 and 4–4). A consequence of these base-pairing requirements is that each strand of a DNA molecule contains a sequence of nucleotides that is exactly **complementary** to the nucleotide sequence of its partner strand.

The Structure of DNA Provides a Mechanism for Heredity

Genes carry biological information that must be copied accurately for transmission to the next generation each time a cell divides to form two daughter cells. Two central biological questions arise from these requirements: how can the information for specifying an organism be carried in chemical form, and how is it accurately copied? The discovery of the structure of the DNA double helix was a landmark in twentieth-century biology because it immediately suggested answers to both questions, thereby resolving at the molecular level the problem of heredity. We discuss briefly the answers to these questions in this section, and we shall examine them in more detail in subsequent chapters.

DNA encodes information through the order, or sequence, of the nucleotides along each strand. Each base—A, C, T, or G—can be considered as a letter in a four-letter alphabet that spells out biological messages in the chemical structure of the DNA. As we saw in Chapter 1, organisms differ from one another because their respective DNA molecules have different nucleotide sequences and, consequently, carry different biological messages. But how is the nucleotide alphabet used to make messages, and what do they spell out?

As discussed above, it was known well before the structure of DNA was determined that genes contain the instructions for producing proteins. The DNA messages must therefore somehow encode proteins (Figure 4–6). This relationship immediately makes the problem easier to understand, because of the chemical character of proteins. As discussed in Chapter 3, the properties of a protein, which are responsible for its biological function, are determined by its three-dimensional structure, and its structure is determined in turn by the linear

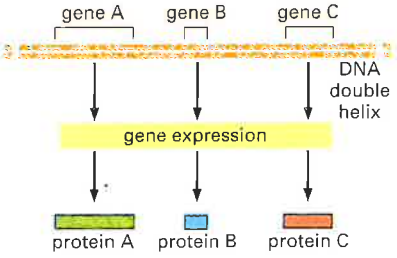


Figure 4–6 The relationship between genetic information carried in DNA and proteins.

Figure 4–7 The nucleotide sequence of the human β-globin gene. This gene carries the information for the amino acid sequence of one of the two types of subunits of the hemoglobin molecule, which carries oxygen in the blood. A different gene, the α-globin gene, carries the information for the other type of hemoglobin subunit (a hemoglobin molecule has four subunits, two of each type). Only one of the two strands of the DNA double helix containing the β-globin gene is shown; the other strand has the exact complementary sequence. By convention, a nucleotide sequence is written from its 5′ end to its 3′ end, and it should be read from left to right in successive lines down the page as though it were normal English text. The DNA sequences highlighted in *yellow* show the three regions of the gene that specify the amino sequence for the β-globin protein. We see in Chapter 6 how the cell connects these three sequences together to synthesize a full-length β-globin protein.

CCCTGTGGAGCCACACCCTAGGGTTGGCCA
ATCTACTCCCAGGAGCAGGGAGGGCAGGAG
CCAGGGCTGGGCATAAAAGTCAGGGCAGAG
CCATCTATTGCTTACATTTGCTTCTGACAC
AACTGTGTTCACTAGCAACTCAAACAGACA
CCATGGTGCACCTGACTCCTGAGGAGAAGT
CTGCCGTTACTGCCCTGTGGGGCAAGGTGA
ACGTGGATGAAGTTGGTGGTGAGGCCCTGG
GCAGGTTGGTATCAAGGTTACAAGACAGGT
TTAAGGAGACCAATAGAAACTGGGCATGTG
GAGACAGAGAAGACTCTTGGGTTTCTGATA
GGCACTGACTCTCTCTGCCTATTGGTCTAT
TTTCCCACCCTTAGGCTGCTGGTGGTCTAC
CCTTGGACCCAGAGGTTCTTTGAGTCCTTT
GGGGATCTGTCCACTCCTGATGCTGTTATG
GGCAACCCTAAGGTGAAGGCTCATGGCAAG
AAAGTGCTCGGTGCCTTTAGTGATGGCCTG
GCTCACCTGGACAACCTCAAGGGCACCTTT
GCCACACTGAGTGAGCTGCACTGTGACAAG
CTGCACGTGGATCCTGAGAACTTCAGGGTG
AGTCTATGGGACCCTTGATGTTTTCTTTCC
CCTTCTTTTCTATGGTTAAGTTCATGTCAT
AGGAAGGGGAGAAGTAACAGGGTACAGTTT
AGAATGGGAAACAGACGAATGATTGCATCA
GTGTGGAAGTCTCAGGATCGTTTTAGTTTC
TTTTATTTGCTGTTCATAACAATTGTTTTC
TTTTGTTTAATTCTTGCTTTCTTTTTTTTT
CTTCTCCGCAATTTTTACTATTATACTTAA
TGCCTTAACATTGTGTATAACAAAAGGAAA
TATCTCTGAGATACATTAAGTAACTTAAAA
AAAAACTTTACACAGTCTGCCTAGTACATT
ACTATTTGGAATATATGTGTGCTTATTTGC
ATATTCATAATCTCCCTACTTTATTTTCTT
TTATTTTTAATTGATACATAATCATTATAC
ATATTTATGGGTTAAAGTGTAATGTTTTAA
TATGTGTACACATATTGACCAAATCAGGGT
AATTTTGCATTTGTAATTTTAAAAAATGCT
TTCTTCTTTTAATATACTTTTTTGTTTATC
TTATTTCTAATACTTTCCCTAATCTCTTTC
TTTCAGGGCAATAATGATACAATGTATCAT
GCCTCTTTGCACCATTCTAAAGAATAACAG
TGATAATTTCTGGGTTAAGGCAATAGCAAT
ATTTCTGCATATAAATATTTCTGCATATAA
ATTGTAACTGATGTAAGAGGTTTCATATTG
CTAATAGCAGCTACAATCCAGCTACCATTC
TGCTTTTATTTTATGGTTGGGATAAGGCTG
GATTATTCTGAGTCCAAGCTAGGCCCTTTT
GCTAATCATGTTCATACCTCTTATCTTCCT
CCCACAGCTCCTGGGCAACGTGCTGGTCTG
TGTGCTGGCCCATCACTTTGGCAAAGAATT
CACCCCACCAGTGCAGGCTGCCTATCAGAA
AGTGGTGGCTGGTGTGGCTAATGCCCTGGC
CCACAAGTATCACTAAGCTCGCTTTCTTGC
TGTCCAATTTCTATTAAAGGTTCCTTTGTT
CCCTAAGTCCAACTACTAAACTGGGGGATA
TTATGAAGGGCCTTGAGCATCTGGATTCTG
CCTAATAAAAAACATTTATTTTCATTGCAA
TGATGTATTTAAATTATTTCTGAATATTTT
ACTAAAAAGGGAATGTGGGAGGTCAGTGCA
TTTAAAACATAAAGAAATGATGAGCTGTTC
AAACCTTGGCAAAATACACTATATCTTAAA
CTCCATGAAAGAAGGTGAGGCTGCAACCAG
CTAATGCACATTGGCAACAGCCCCTGATGC
CTATGCCTTATTCATCCCTCAGAAAAGGAT
TCTTGTAGAGGCTTGATTTGCAGGTTAAAG
TTTTGCTATGCTGTATTTTACATTACTTAT
TGTTTTAGCTGTCCTCATGAATGTCTTTTC

sequence of the amino acids of which it is composed. The linear sequence of nucleotides in a gene must therefore somehow spell out the linear sequence of amino acids in a protein. The exact correspondence between the four-letter nucleotide alphabet of DNA and the twenty-letter amino acid alphabet of proteins—the genetic code—is not obvious from the DNA structure, and it took over a decade after the discovery of the double helix before it was worked out. In Chapter 6 we describe this code in detail in the course of elaborating the process, known as *gene expression*, through which a cell translates the nucleotide sequence of a gene into the amino acid sequence of a protein.

The complete set of information in an organism's DNA is called its **genome**, and it carries the information for all the proteins the organism will ever synthesize. (The term genome is also used to describe the DNA that carries this information.) The amount of information contained in genomes is staggering: for example, a typical human cell contains 2 meters of DNA. Written out in the four-letter nucleotide alphabet, the nucleotide sequence of a very small human gene occupies a quarter of a page of text (Figure 4–7), while the complete sequence of nucleotides in the human genome would fill more than a thousand books the size of this one. In addition to other critical information, it carries the instructions for about 30,000 distinct proteins.

At each cell division, the cell must copy its genome to pass it to both daughter cells. The discovery of the structure of DNA also revealed the principle that makes this copying possible: because each strand of DNA contains a sequence of nucleotides that is exactly complementary to the nucleotide sequence of its partner strand, each strand can act as a **template**, or mold, for the synthesis of a new complementary strand. In other words, if we designate the two DNA strands as S and S′, strand S can serve as a template for making a new strand S′, while strand S′ can serve as a template for making a new strand S (Figure 4–8).

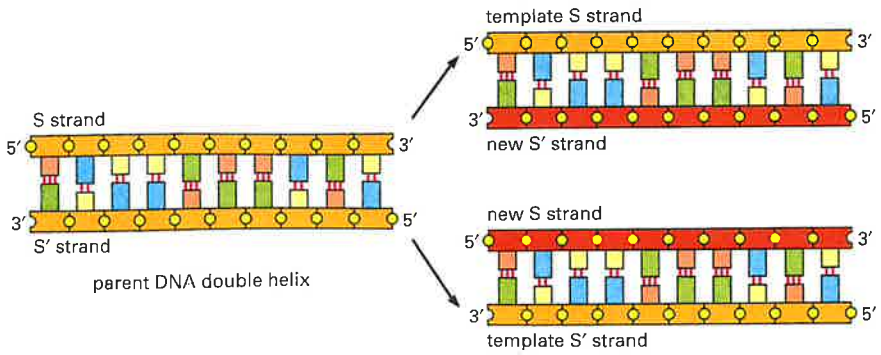


Figure 4–8 DNA as a template for its own duplication. As the nucleotide A successfully pairs only with T, and G with C, each strand of DNA can specify the sequence of nucleotides in its complementary strand. In this way, double-helical DNA can be copied precisely.

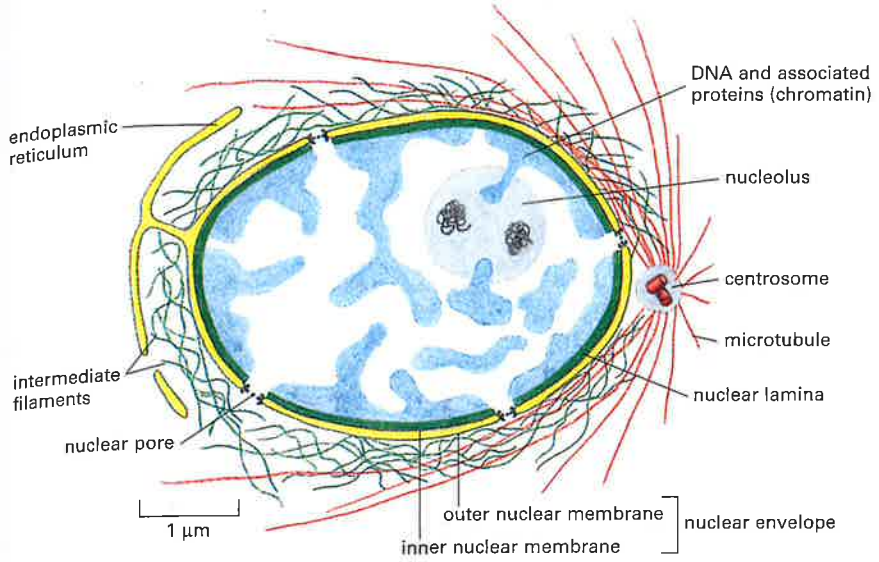


Figure 4–9 A cross-sectional view of a typical cell nucleus. The nuclear envelope consists of two membranes, the outer one being continuous with the endoplasmic reticulum membrane (see also Figure 12–9). The space inside the endoplasmic reticulum (the ER lumen) is colored *yellow;* it is continuous with the space between the two nuclear membranes. The lipid bilayers of the inner and outer nuclear membranes are connected at each nuclear pore. Two networks of intermediate filaments *(green)* provide mechanical support for the nuclear envelope; the intermediate filaments inside the nucleus form a special supporting structure called the nuclear lamina.

Thus, the genetic information in DNA can be accurately copied by the beautifully simple process in which strand S separates from strand S′, and each separated strand then serves as a template for the production of a new complementary partner strand that is identical to its former partner.

The ability of each strand of a DNA molecule to act as a template for producing a complementary strand enables a cell to copy, or *replicate,* its genes before passing them on to its descendants. In the next chapter we describe the elegant machinery the cell uses to perform this enormous task.

In Eucaryotes, DNA Is Enclosed in a Cell Nucleus

Nearly all the DNA in a eucaryotic cell is sequestered in a nucleus, which occupies about 10% of the total cell volume. This compartment is delimited by a *nuclear envelope* formed by two concentric lipid bilayer membranes that are punctured at intervals by large nuclear pores, which transport molecules between the nucleus and the cytosol. The nuclear envelope is directly connected to the extensive membranes of the endoplasmic reticulum. It is mechanically supported by two networks of intermediate filaments: one, called the *nuclear lamina,* forms a thin sheetlike meshwork inside the nucleus, just beneath the inner nuclear membrane; the other surrounds the outer nuclear membrane and is less regularly organized (Figure 4–9).

The nuclear envelope allows the many proteins that act on DNA to be concentrated where they are needed in the cell, and, as we see in subsequent chapters, it also keeps nuclear and cytosolic enzymes separate, a feature that is crucial for the proper functioning of eucaryotic cells. Compartmentalization, of which the nucleus is an example, is an important principle of biology; it serves to establish an environment in which biochemical reactions are facilitated by the high concentration of both substrates and the enzymes that act on them.

Summary

Genetic information is carried in the linear sequence of nucleotides in DNA. Each molecule of DNA is a double helix formed from two complementary strands of nucleotides held together by hydrogen bonds between G-C and A-T base pairs. Duplication of the genetic information occurs by the use of one DNA strand as a template for formation of a complementary strand. The genetic information stored in an organism's DNA contains the instructions for all the proteins the organism will ever synthesize. In eucaryotes, DNA is contained in the cell nucleus.

CHROMOSOMAL DNA AND ITS PACKAGING IN THE CHROMATIN FIBER

The most important function of DNA is to carry genes, the information that specifies all the proteins that make up an organism—including information about when, in what types of cells, and in what quantity each protein is to be made. The genomes of eucaryotes are divided up into chromosomes, and in this section we see how genes are typically arranged on each chromosome. In addition, we describe the specialized DNA sequences that allow a chromosome to be accurately duplicated and passed on from one generation to the next.

We also confront the serious challenge of DNA packaging. Each human cell contains approximately 2 meters of DNA if stretched end-to-end; yet the nucleus of a human cell, which contains the DNA, is only about 6 μm in diameter. This is geometrically equivalent to packing 40 km (24 miles) of extremely fine thread into a tennis ball! The complex task of packaging DNA is accomplished by specialized proteins that bind to and fold the DNA, generating a series of coils and loops that provide increasingly higher levels of organization, preventing the DNA from becoming an unmanageable tangle. Amazingly, although the DNA is very tightly folded, it is compacted in a way that allows it to easily become available to the many enzymes in the cell that replicate it, repair it, and use its genes to produce proteins.

Eucaryotic DNA Is Packaged into a Set of Chromosomes

In eucaryotes, the DNA in the nucleus is divided between a set of different **chromosomes**. For example, the human genome—approximately 3.2×10^9 nucleotides—is distributed over 24 different chromosomes. Each chromosome consists of a single, enormously long linear DNA molecule associated with proteins that fold and pack the fine DNA thread into a more compact structure. The complex of DNA and protein is called *chromatin* (from the Greek *chroma*, "color," because of its staining properties). In addition to the proteins involved in packaging the DNA, chromosomes are also associated with many proteins required for the processes of gene expression, DNA replication, and DNA repair.

Bacteria carry their genes on a single DNA molecule, which is usually circular (see Figure 1–30). This DNA is associated with proteins that package and condense the DNA, but they are different from the proteins that perform these functions in eucaryotes. Although often called the bacterial "chromosome," it does not have the same structure as eucaryotic chromosomes, and less is known about how the bacterial DNA is packaged. Even less is known about how DNA is compacted in archaea. Therefore, our discussion of chromosome structure will focus almost entirely on eucaryotic chromosomes.

With the exception of the germ cells, and a few highly specialized cell types that cannot multiply and lack DNA altogether (for example, red blood cells), each human cell contains two copies of each chromosome, one inherited from the mother and one from the father. The maternal and paternal chromosomes of a pair are called **homologous chromosomes (homologs)**. The only nonhomologous chromosome pairs are the sex chromosomes in males, where a *Y chromosome* is inherited from the father and an *X chromosome* from the mother. Thus, each human cell contains a total of 46 chromosomes—22 pairs common to both males and females, plus two so-called sex chromosomes (X and Y in males, two Xs in females). *DNA hybridization* (described in detail in Chapter 8) can be used to distinguish these human chromosomes by "painting" each one a different color (Figure 4–10). Chromosome painting is typically done at the stage in the cell cycle when chromosomes are especially compacted and easy to visualize (mitosis, see below).

Another more traditional way to distinguish one chromosome from another is to stain them with dyes that produce a striking and reliable pattern of bands along each mitotic chromosome (Figure 4–11). The structural bases for these banding patterns are not well understood, and we return to this issue at the end of the chapter. Nevertheless, the pattern of bands on each type of chromosome is unique, allowing each chromosome to be identified and numbered.

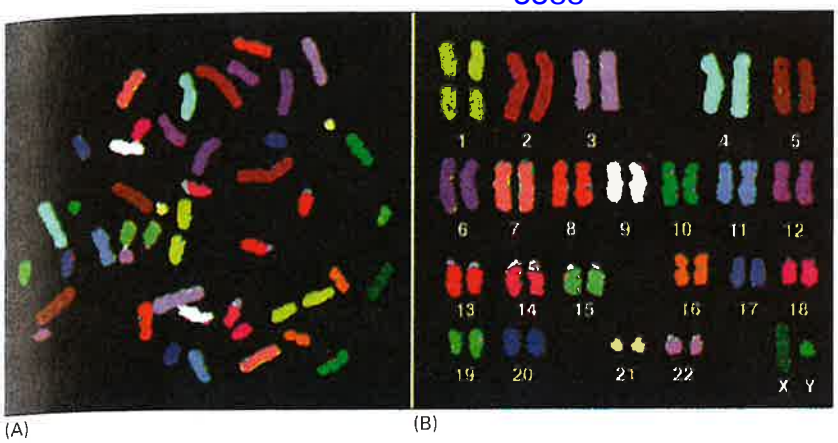


Figure 4–10 Human chromosomes. These chromosomes, from a male, were isolated from a cell undergoing nuclear division (mitosis) and are therefore highly compacted. Each chromosome has been "painted" a different color to permit its unambiguous identification under the light microscope. Chromosome painting is performed by exposing the chromosomes to a collection of human DNA molecules that have been coupled to a combination of fluorescent dyes. For example, DNA molecules derived from chromosome 1 are labeled with one specific dye combination, those from chromosome 2 with another, and so on. Because the labeled DNA can form base pairs, or hybridize, only to the chromosome from which it was derived (discussed in Chapter 8), each chromosome is differently labeled. For such experiments, the chromosomes are subjected to treatments that separate the double-helical DNA into individual strands, designed to permit base-pairing with the single-stranded labeled DNA while keeping the chromosome structure relatively intact. (A) The chromosomes visualized as they originally spilled from the lysed cell. (B) The same chromosomes artificially lined up in their numerical order. This arrangement of the full chromosome set is called a karyotype. (From E. Schröck et al., *Science* 273:494–497, 1996. © AAAS.)

The display of the 46 human chromosomes at mitosis is called the human **karyotype**. If parts of chromosomes are lost, or switched between chromosomes, these changes can be detected by changes in the banding patterns or by changes in the pattern of chromosome painting (Figure 4–12). Cytogeneticists use these alterations to detect chromosome abnormalities that are associated with inherited defects or with certain types of cancer that arise through the rearrangement of chromosomes in somatic cells.

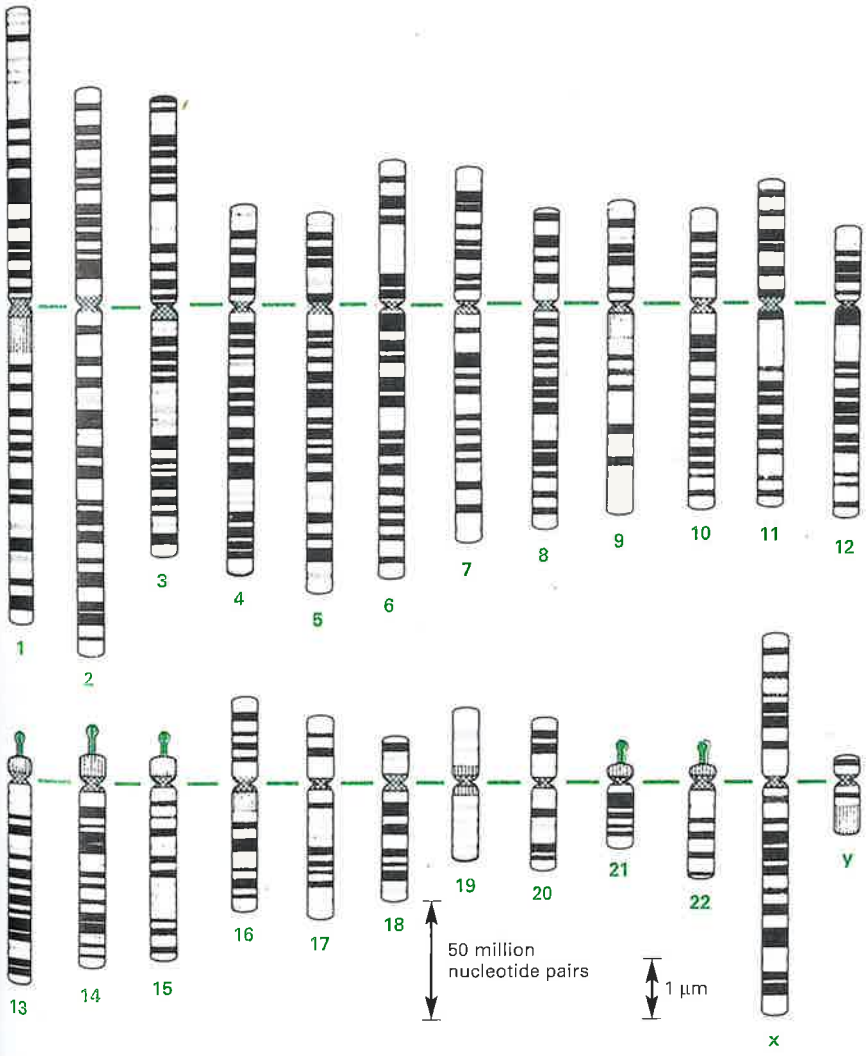


Figure 4–11 The banding patterns of human chromosomes. Chromosomes 1–22 are numbered in approximate order of size. A typical human somatic (non-germ line) cell contains two of each of these chromosomes, plus two sex chromosomes—two X chromosomes in a female, one X and one Y chromosome in a male. The chromosomes used to make these maps were stained at an early stage in mitosis, when the chromosomes are incompletely compacted. The *horizontal green line* represents the position of the centromere (see Figure 4–22), which appears as a constriction on mitotic chromosomes; the knobs on chromosomes 13, 14, 15, 21, and 22 indicate the positions of genes that code for the large ribosomal RNAs (discussed in Chapter 6). These patterns are obtained by staining chromosomes with Giemsa stain, and they can be observed under the light microscope. (Adapted from U. Franke, *Cytogenet. Cell Genet.* 31:24–32, 1981.)

Chromosomes Contain Long Strings of Genes

The most important function of chromosomes is to carry genes—the functional units of heredity. A gene is usually defined as a segment of DNA that contains the instructions for making a particular protein (or a set of closely related proteins). Although this definition holds for the majority of genes, several percent of genes produce an RNA molecule, instead of a protein, as their final product. Like proteins, these RNA molecules perform a diverse set of structural and catalytic functions in the cell, and we discuss them in detail in subsequent chapters.

As might be expected, a correlation exists between the complexity of an organism and the number of genes in its genome (see Table 1–1). For example, total gene numbers range from less than 500 for simple bacteria to about 30,000 for humans. Bacteria and some single-celled eucaryotes have especially compact genomes; the complete nucleotide sequence of their genomes reveals that the DNA molecules that make up their chromosomes are little more than strings of closely packed genes (Figure 4–13; see also Figure 1–30). However, chromosomes from many eucaryotes (including humans) contain, in addition to genes, a large excess of interspersed DNA that does not seem to carry critical information. Sometimes called junk DNA to signify that its usefulness to the cell has not been demonstrated, the particular nucleotide sequence of this DNA may not be important; but the DNA itself, by acting as spacer material, may be crucial for the long-term evolution of the species and for the proper expression of genes. These issues are taken up in detail in Chapter 7.

In general, the more complex the organism, the larger its genome, but because of differences in the amount of excess DNA, the relationship is not systematic (see Figure 1–38). For example, the human genome is 200 times larger than that of the yeast *S. cerevisiae*, but 30 times smaller than that of some plants and amphibians and 200 times smaller than a species of amoeba. Moreover, because of differences in the amount of excess DNA, the genomes of similar organisms (bony fish, for example) can vary several hundredfold in their DNA content, even though they contain roughly the same number of genes. Whatever the excess DNA may do, it seems clear that it is not a great handicap for a higher eucaryotic cell to carry a large amount of it.

The apportionment of the genome over chromosomes also differs from one eucaryotic species to the next. For example, compared with 46 for humans,

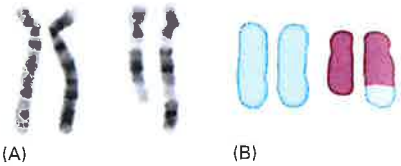


Figure 4–12 An aberrant human chromosome. (A) Two pairs of chromosomes, stained with Giemsa (see Figure 4–11), from a patient with ataxia, a disease characterized by progressive deterioration of motor skills. The patient has a normal pair of chromosome 4s *(left-hand pair)*, but one normal chromosome 12 and one aberrant chromosome 12, as seen by its greater length *(right-hand pair)*. The additional material contained on the aberrant chromosome 12 was deduced, from its pattern of bands, as a piece of chromosome 4 that had become attached to chromosome 12 through an abnormal recombination event, called a chromosomal translocation. (B) The same two chromosome pairs, "painted" *blue* for chromosome 4 DNA and *purple* for chromosome 12 DNA. The two techniques give rise to the same conclusion regarding the nature of the aberrant chromosome 12, but chromosome painting provides better resolution, and the clear identification of even short pieces of chromosomes that have become translocated. However, Giemsa staining is easier to perform. (From E. Schröck et al., *Science* 273:494–497, 1996. © AAAS.)

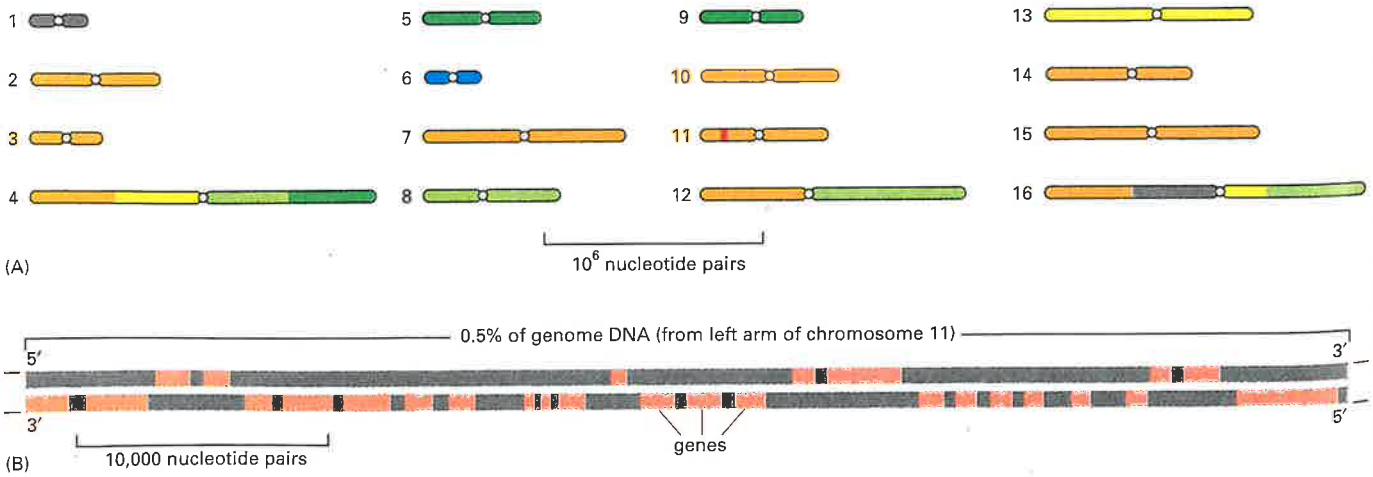


Figure 4–13 The genome of *S. cerevisiae* (budding yeast). (A) The genome is distributed over 16 chromosomes, and its complete nucleotide sequence was determined by a cooperative effort involving scientists working in many different locations, as indicated *(gray,* Canada; *orange,* European Union; *yellow,* United Kingdom; *blue,* Japan; *light green,* St Louis, Missouri; *dark green,* Stanford, California). The constriction present on each chromosome represents the position of its centromere (see Figure 4–22). (B) A small region of chromosome 11, highlighted in *red* in part A, is magnified to show the high density of genes characteristic of this species. As indicated by *orange,* some genes are transcribed from the lower strand (see Figure 1–5), while others are transcribed from the upper strand. There are about 6000 genes in the complete genome, which is 12,147,813 nucleotide pairs long.

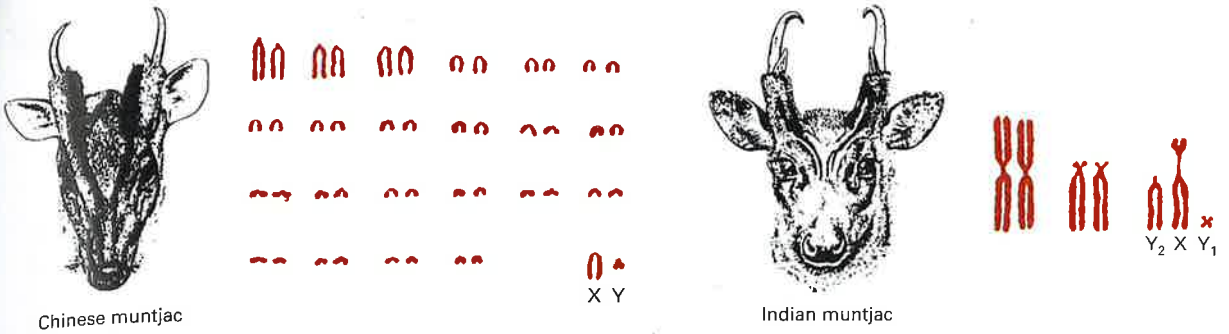


Figure 4–14 Two closely related species of deer with very different chromosome numbers. In the evolution of the Indian muntjac, initially separate chromosomes fused, without having a major effect on the animal. These two species have roughly the same number of genes. (Adapted from M.W. Strickberger, Evolution, 3rd edition, 2000, Sudbury, MA: Jones & Bartlett Publishers.)

somatic cells from a species of small deer contain only 6 chromosomes, while those from a species of carp contain over 100. Even closely related species with similar genome sizes can have very different numbers and sizes of chromosomes (Figure 4–14). Thus, there is no simple relationship between chromosome number, species complexity, and total genome size. Rather, the genomes and chromosomes of modern-day species have each been shaped by a unique history of seemingly random genetic events, acted on by selection pressures.

The Nucleotide Sequence of the Human Genome Shows How Genes Are Arranged in Humans

When the DNA sequence of human chromosome 22, one of the smallest human chromosomes (see Figure 4–11), was completed in 1999, it became possible for the first time to see exactly how genes are arranged along an entire vertebrate chromosome (Figure 4–15 and Table 4–1). With the publication of the "first draft" of the entire human genome in 2001, the genetic landscape of all human chromosomes suddenly came into sharp focus. The sheer quantity of information

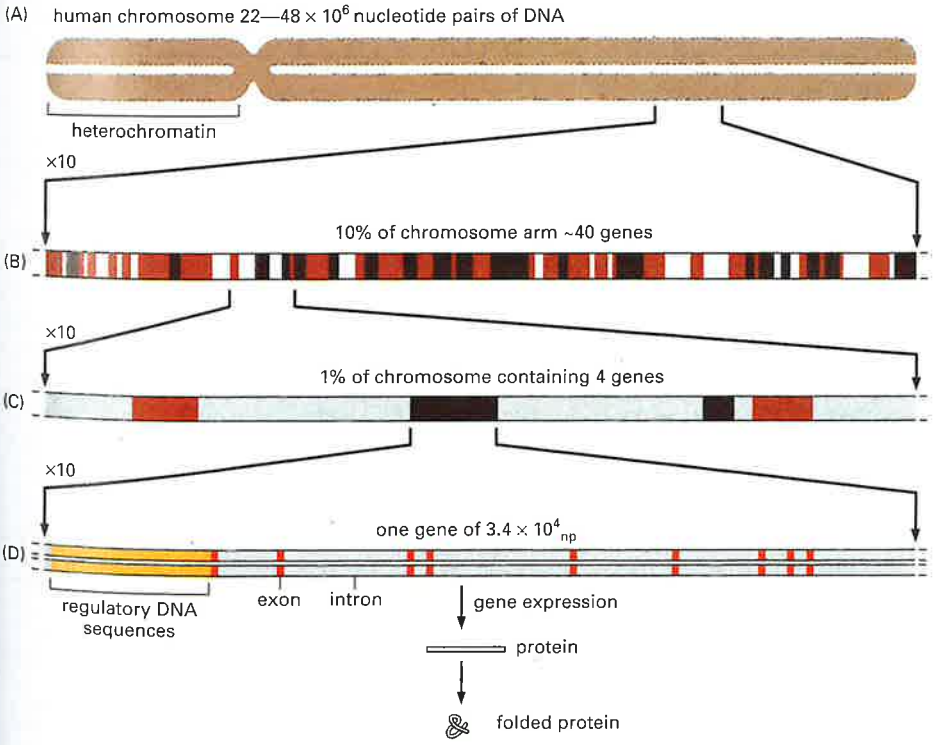


Figure 4–15 The organization of genes on a human chromosome. (A) Chromosome 22, one of the smallest human chromosomes, contains 48×10^6 nucleotide pairs and makes up approximately 1.5% of the entire human genome. Most of the left arm of chromosome 22 consists of short repeated sequences of DNA that are packaged in a particularly compact form of chromatin (heterochromatin), which is discussed later in this chapter. (B) A tenfold expansion of a portion of chromosome 22, with about 40 genes indicated. Those in *dark brown* are known genes and those in *light brown* are predicted genes. (C) An expanded portion of (B) shows the entire length of several genes. (D) The intron–exon arrangement of a typical gene is shown after a further tenfold expansion. Each exon *(red)* codes for a portion of the protein, while the DNA sequence of the introns *(gray)* is relatively unimportant. The entire human genome (3.2×10^9 nucleotide pairs) is distributed over 22 autosomes and 2 sex chromosomes (see Figures 4–10 and 4–11). The term *human genome sequence* refers to the complete nucleotide sequence of DNA in these 24 chromosomes. Being diploid, a human somatic cell therefore contains roughly twice this amount of DNA. Humans differ from one another by an average of one nucleotide in every thousand, and a wide variety of humans contributed DNA for the genome sequencing project. The published human genome sequence is therefore a composite of many individual sequences. (Adapted from International Human Genome Sequencing Consortium, *Nature* 409:860–921, 2001.)

provided by the Human Genome Project is unprecedented in biology (Figure 4–16 and Table 4–1); the human genome is 25 times larger than any other genome sequenced so far, and is 8 times as large as the sum of all previously sequenced genomes. At its peak, the Human Genome Project generated raw nucleotide sequences at a rate of 1000 nucleotides per second around the clock. It will be many decades before this information is fully analyzed, but it will continue to stimulate many new experiments and has already affected the content of all the chapters in this book.

Although there are many aspects to analyzing the human genome, here we simply make a few generalizations regarding the arrangement of genes in human chromosomes. The first striking feature of the human genome is how little of it (only a few percent) codes for proteins or structural and catalytic RNAs (Figure 4–17). Much of the remaining chromosomal DNA is made up of short, mobile pieces of DNA that have gradually inserted themselves in the chromosome over evolutionary time. We discuss these *transposable elements* in detail in later chapters.

A second notable feature of the human genome is the large average gene size of 27,000 nucleotide pairs. As discussed above, a typical gene carries in its linear sequence of nucleotides the information for the linear sequence of the amino acids of a protein. Only about 1300 nucleotide pairs are required to encode a protein of average size (about 430 amino acids in humans). Most of the remaining DNA in a gene consists of long stretches of noncoding DNA that interrupt the relatively short segments of DNA that code for protein. The coding sequences are called *exons*; the intervening (noncoding) sequences are called *introns* (see Figure 4–15 and Table 4–1).

The majority of human genes thus consist of a long string of alternating exons and introns, with most of the gene consisting of introns. In contrast, the majority of genes from organisms with compact genomes lack introns. This accounts for the much smaller size of their genes (about one-twentieth that of human genes), as well as for the much higher fraction of coding DNA in their chromosomes. In addition to introns and exons, each gene is associated with *regulatory DNA sequences*, which are responsible for ensuring that the gene is

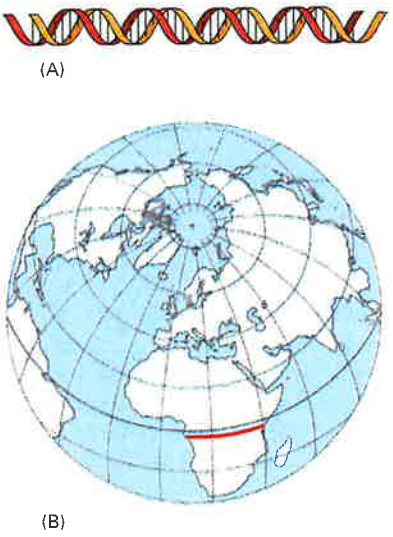


Figure 4–16 Scale of the human genome. If each nucleotide pair is drawn as 1 mm as in (A), then the human genome would extend 3200 km (approximately 2000 miles), far enough to stretch across the center of Africa, the site of our human origins (*red line* in B). At this scale, there would be, on average, a protein-coding gene every 300 m. An average gene would extend for 30 m, but the coding sequences in this gene would add up to only just over a meter.

TABLE 4–1 Vital Statistics of Human Chromosome 22 and the Entire Human Genome

	CHROMOSOME 22	HUMAN GENOME
DNA length	48×10^6 nucleotide pairs*	3.2×10^9
Number of genes	approximately 700	approximately 30,000
Smallest protein-coding gene	1000 nucleotide pairs	not analyzed
Largest gene	583,000 nucleotide pairs	2.4×10^6 nucleotide pairs
Mean gene size	19,000 nucleotide pairs	27,000 nucleotide pairs
Smallest number of exons per gene	1	1
Largest number of exons per gene	54	178
Mean number of exons per gene	5.4	8.8
Smallest exon size	8 nucleotide pairs	not analyzed
Largest exon size	7600 nucleotide pairs	17,106 nucleotide pairs
Mean exon size	266 nucleotide pairs	145 nucleotide pairs
Number of pseudogenes**	more than 134	not analyzed
Percentage of DNA sequence in exons (protein coding sequences)	3%	1.5%
Percentage of DNA in high-copy repetitive elements	42%	approximately 50%
Percentage of total human genome	1.5%	100%

* The nucleotide sequence of 33.8×10^6 nucleotides is known; the rest of the chromosome consists primarily of very short repeated sequences that do not code for proteins or RNA.

** A pseudogene is a nucleotide sequence of DNA closely resembling that of a functional gene, but containing numerous deletion mutations that prevent its proper expression. Most pseudogenes arise from the duplication of a functional gene followed by the accumulation of damaging mutations in one copy.

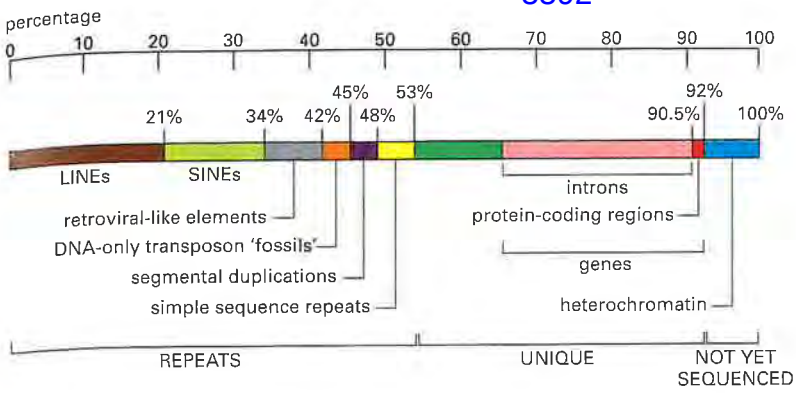


Figure 4–17 Representation of the nucleotide sequence content of the human genome. LINES, SINES, retroviral-like elements, and DNA-only transposons are all mobile genetic elements that have multiplied in our genome by replicating themselves and inserting the new copies in different positions. Mobile genetic elements are discussed in Chapter 5. Simple sequence repeats are short nucleotide sequences (less than 14 nucleotide pairs) that are repeated again and again for long stretches. Segmental duplications are large blocks of the genome (1000–200,000 nucleotide pairs) that are present at two or more locations in the genome. Over half of the unique sequence consists of genes and the remainder is probably regulatory DNA. Most of the DNA present in heterochromatin, a specialized type of chromatin (discussed later in this chapter) that contains relatively few genes, has not yet been sequenced. (Adapted from Unveiling the Human Genome, Supplement to the Wellcome Trust Newsletter. London: Wellcome Trust, February 2001.)

expressed at the proper level and time, and the proper type of cell. In humans, the regulatory sequences for a typical gene are spread out over tens of thousands of nucleotide pairs. As would be expected, these regulatory sequences are more compressed in organisms with compact genomes. We discuss in Chapter 7 how regulatory DNA sequences work.

Finally, the nucleotide sequence of the human genome has revealed that the critical information seems to be in an alarming state of disarray. As one commentator described our genome, "In some ways it may resemble your garage/bedroom/refrigerator/life: highly individualistic, but unkempt; little evidence of organization; much accumulated clutter (referred to by the uninitiated as 'junk'); virtually nothing ever discarded; and the few patently valuable items indiscriminately, apparently carelessly, scattered throughout."

Comparisons Between the DNAs of Related Organisms Distinguish Conserved and Nonconserved Regions of DNA Sequence

A major obstacle in interpreting the nucleotide sequences of human chromosomes is the fact that much of the sequence is probably unimportant. Moreover, the coding regions of the genome (the exons) are typically found in short segments (average size about 145 nucleotide pairs) floating in a sea of DNA whose exact nucleotide sequence is of little consequence. This arrangement makes it very difficult to identify all the exons in a stretch of DNA sequence; even harder is the determination of where a gene begins and ends and how many exons it spans. Accurate gene identification requires approaches that extract information from the inherently low signal-to-noise ratio of the human genome, and we describe some of them in Chapter 8. Here we discuss the most general approach, one that has the potential to identify not only coding sequences but also additional DNA sequences that are important. It is based on the observation that sequences that have a function are conserved during evolution, whereas those without a function are free to mutate randomly. The strategy is therefore to compare the human sequence with that of the corresponding regions of a related genome, such as that of the mouse. Humans and mice are thought to have diverged from a common mammalian ancestor about 100×10^6 years ago, which is long enough for the majority of nucleotides in their genomes to have been changed by random mutational events. Consequently, the only regions that will have remained closely similar in the two genomes are those in which mutations would have impaired function and put the animals carrying them at a disadvantage, resulting in their elimination from the population by natural selection. Such closely similar regions are known as *conserved regions*. In general, conserved regions represent functionally important exons and regulatory sequences. In contrast, *nonconserved regions* represent DNA whose sequence is generally not critical for function. By revealing in this way the results of a very long natural "experiment," comparative DNA sequencing studies highlight the most interesting regions in genomes.

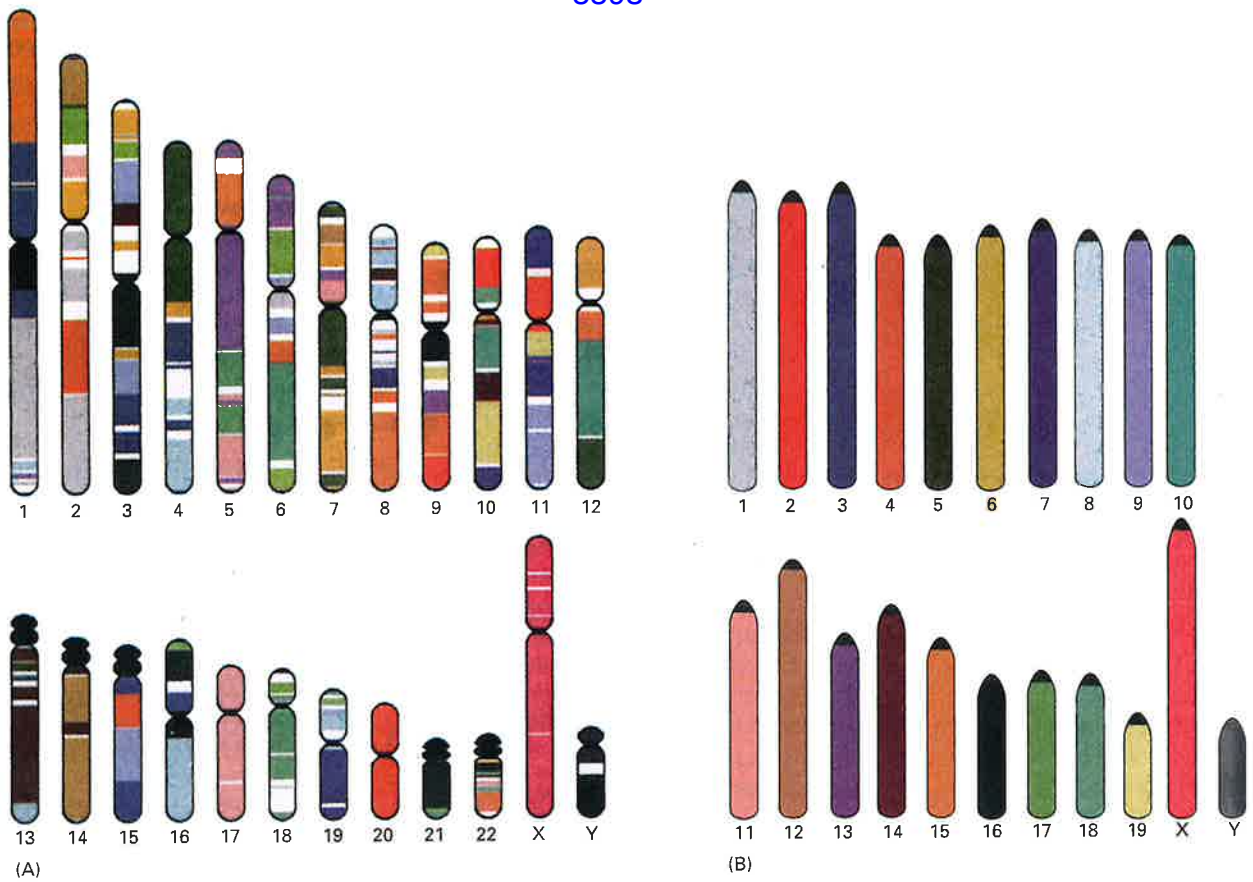


Figure 4–18 Conserved synteny between the human and mouse genomes. Regions from different mouse chromosomes (indicated by the colors of each mouse in B) show conserved synteny (gene order) with the indicated regions of the human genome (A). For example the genes present in the upper portion of human chromosome 1 *(orange)* are present in the same order in a portion of mouse chromosome 4. Regions of human chromosomes that are composed primarily of short, repeated sequences are shown in *black*. Mouse centromeres (indicated in *black* in B) are located at the ends of chromosomes; no known genes lie beyond the centromere on any mouse chromosome. For the most part, human centromeres, indicated by constrictions, occupy more internal positions on chromosomes (see Figure 4–11). (Adapted from International Human Genome Sequencing Consortium, *Nature* 409:860–921, 2001.)

Comparative studies of this kind have revealed not only that mice and humans share most of the same genes, but also that large blocks of the mouse and human genomes contain these genes in the same order, a feature called *conserved synteny* (Figure 4–18). Conserved synteny can also be revealed by chromosome painting, and this technique has been used to reconstruct the evolutionary history of our own chromosomes by comparing them with those from other mammals (Figure 4–19).

Chromosomes Exist in Different States Throughout the Life of a Cell

We have seen how genes are arranged in chromosomes, but to form a functional chromosome, a DNA molecule must be able to do more than simply carry genes: it must be able to replicate, and the replicated copies must be separated and reliably partitioned into daughter cells at each cell division. This process occurs through an ordered series of stages, collectively known as the **cell cycle**. The cell cycle is briefly summarized in Figure 4–20, and discussed in detail in Chapter 17. Only two of the stages of the cycle concern us in this chapter. During *interphase* chromosomes are replicated, and during *mitosis* they become highly condensed and then are separated and distributed to the two daughter nuclei. The highly condensed chromosomes in a dividing cell are known as *mitotic chromosomes*. This is the form in which chromosomes are most easily visualized; in fact, all the images of chromosomes shown so far in the chapter are of chromosomes in mitosis. This condensed state is important in allowing the duplicated chromosomes to be separated by the mitotic spindle during cell division, as discussed in Chapter 18.

During the portions of the cell cycle when the cell is not dividing, the chromosomes are extended and much of their chromatin exists as long, thin tangled

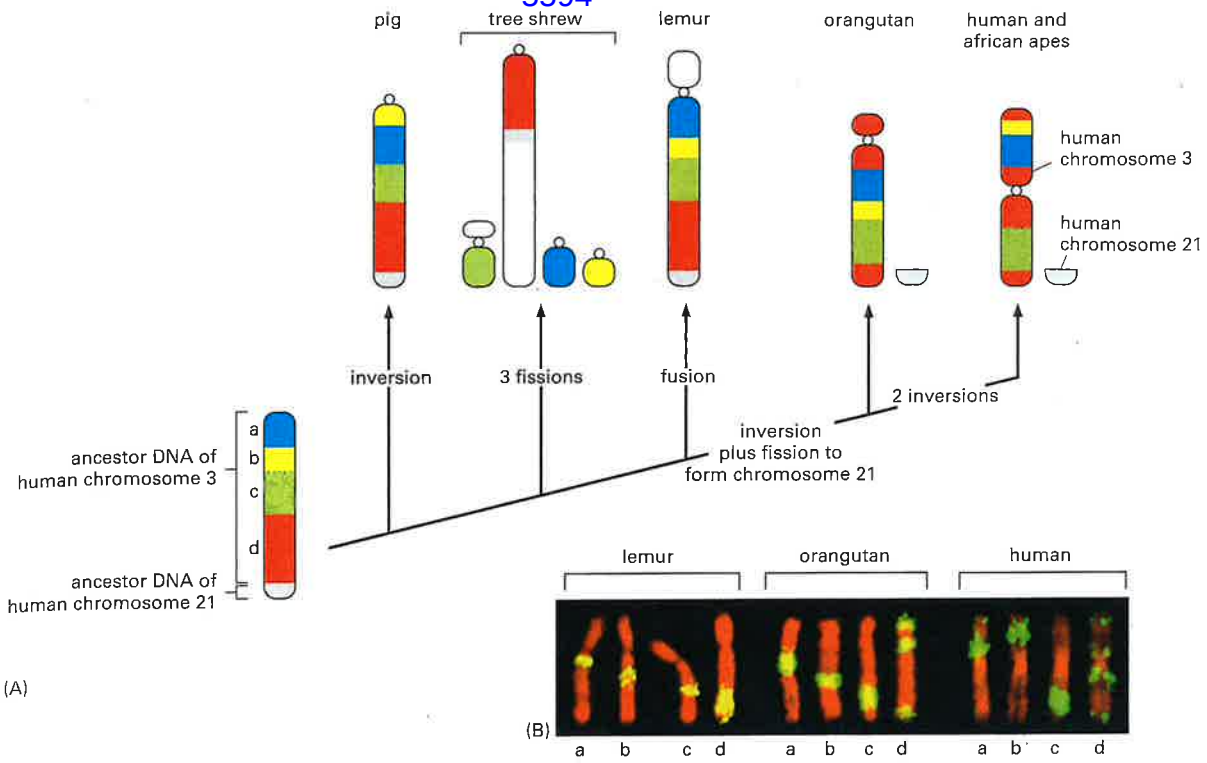


Figure 4–19 A proposed evolutionary history of human chromosome 3 and its relatives in other mammals. (A) At the *lower left* is the order of chromosome 3 segments hypothesized to be present on a chromosome of a mammalian ancestor. Along the *top* are the patterns of chromosome sequences found in the chromosomes of modern mammals. The minimum changes necessary to account for the appearance of the modern chromosomes from the hypothetical ancestor are marked along each branch. In mammals, these types of changes in chromosome organization are thought to occur once every 5–10 × 10^6 years. The *small circles* depicted in the modern chromosomes represent the positions of centromeres. (B) Some of the chromosome painting experiments that led to the diagram in (A). Each image shows the chromosome most closely related to human chromosome 3, painted *green* by hybridization with different segments of DNA, lettered a, b, c, and d along the *bottom* of the figure. These letters correspond to the colored segments of the diagram in (A). (From S. Müller et al., *Proc. Natl. Acad. Sci. USA* 97:206–211, 2000. © National Academy of Sciences.)

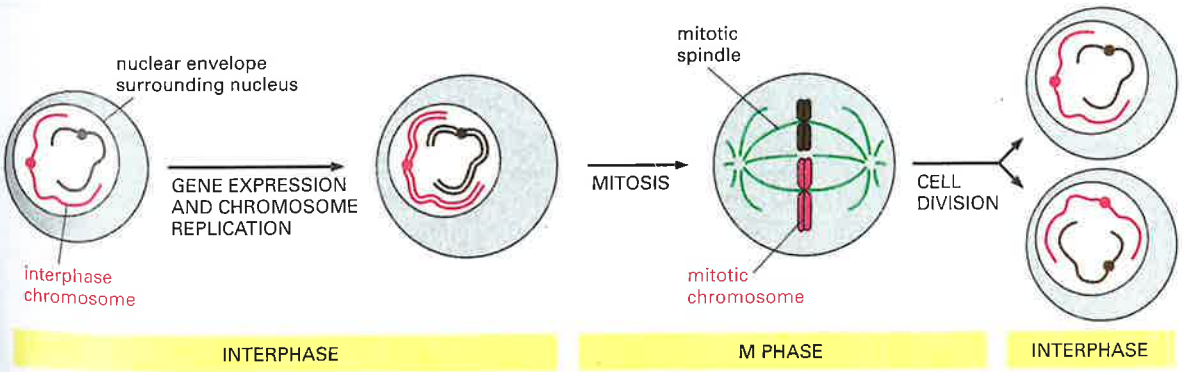


Figure 4–20 A simplified view of the eucaryotic cell cycle. During interphase, the cell is actively expressing its genes and is therefore synthesizing proteins. Also, during interphase and before cell division, the DNA is replicated and the chromosomes are duplicated. Once DNA replication is complete, the cell can enter *M phase*, when mitosis occurs and the nucleus is divided into two daughter nuclei. During this stage, the chromosomes condense, the nuclear envelope breaks down, and the mitotic spindle forms from microtubules and other proteins. The condensed mitotic chromosomes are captured by the mitotic spindle, and one complete set of chromosomes is then pulled to each end of the cell. A nuclear envelope re-forms around each chromosome set, and in the final step of M phase, the cell divides to produce two daughter cells. Most of the time in the cell cycle is spent in interphase; M phase is brief in comparison, occupying only about an hour in many mammalian cells.

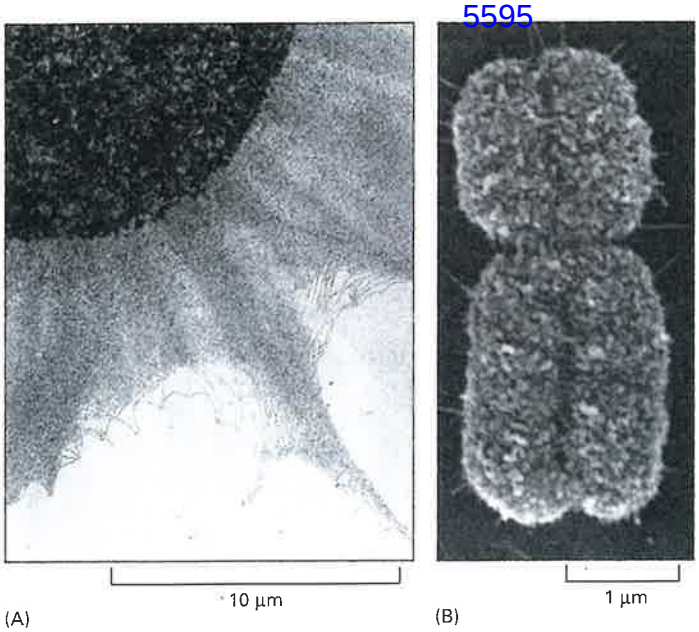


Figure 4–21 A comparison of extended interphase chromatin with the chromatin in a mitotic chromosome. (A) An electron micrograph showing an enormous tangle of chromatin spilling out of a lysed interphase nucleus. (B) A scanning electron micrograph of a mitotic chromosome: a condensed duplicated chromosome in which the two new chromosomes are still linked together (see Figure 4–22). The constricted region indicates the position of the centromere. Note the difference in scales. (A, courtesy of Victoria Foe; B, courtesy of Terry D. Allen.)

threads in the nucleus so that individual chromosomes cannot be easily distinguished (Figure 4–21). We refer to chromosomes in this extended state as *interphase chromosomes*.

Each DNA Molecule That Forms a Linear Chromosome Must Contain a Centromere, Two Telomeres, and Replication Origins

A chromosome operates as a distinct structural unit: for a copy to be passed on to each daughter cell at division, each chromosome must be able to replicate, and the newly replicated copies must subsequently be separated and partitioned correctly into the two daughter cells. These basic functions are controlled by three types of specialized nucleotide sequence in the DNA, each of which binds specific proteins that guide the machinery that replicates and segregates chromosomes (Figure 4–22).

Experiments in yeasts, whose chromosomes are relatively small and easy to manipulate, have identified the minimal DNA sequence elements responsible for each of these functions. One type of nucleotide sequence acts as a DNA **replication origin**, the location at which duplication of the DNA begins. Eucaryotic chromosomes contain many origins of replication to ensure that the entire chromosome can be replicated rapidly, as discussed in detail in Chapter 5.

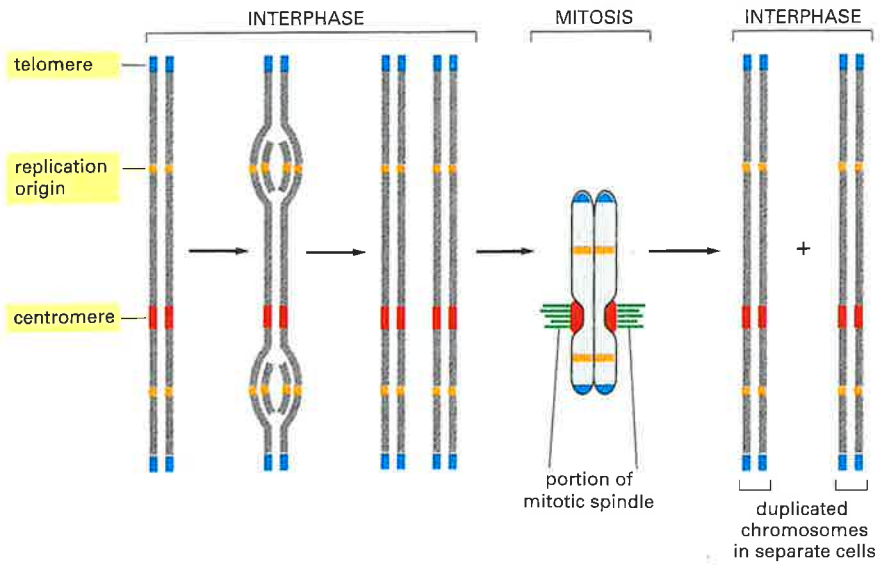


Figure 4–22 The three DNA sequences required to produce a eucaryotic chromosome that can be replicated and then segregated at mitosis. Each chromosome has multiple origins of replication, one centromere, and two telomeres. Shown here is the sequence of events a typical chromosome follows during the cell cycle. The DNA replicates in interphase beginning at the origins of replication and proceeding bidirectionally from the origins across the chromosome. In M phase, the centromere attaches the duplicated chromosomes to the mitotic spindle so that one copy is distributed to each daughter cell during mitosis. The centromere also helps to hold the duplicated chromosomes together until they are ready to be moved apart. The telomeres form special caps at each chromosome end.

After replication, the two daughter chromosomes remain attached to one another and, as the cell cycle proceeds, are condensed further to produce mitotic chromosomes. The presence of a second specialized DNA sequence, called a **centromere**, allows one copy of each duplicated and condensed chromosome to be pulled into each daughter cell when a cell divides. A protein complex called a *kinetochore* forms at the centromere and attaches the duplicated chromosomes to the mitotic spindle, allowing them to be pulled apart (discussed in Chapter 18).

The third specialized DNA sequence forms **telomeres**, the ends of a chromosome. Telomeres contain repeated nucleotide sequences that enable the ends of chromosomes to be efficiently replicated. Telomeres also perform another function: the repeated telomere DNA sequences, together with the regions adjoining them, form structures that protect the end of the chromosome from being recognized by the cell as a broken DNA molecule in need of repair. We discuss this type of repair and the other features of telomeres in Chapter 5.

In yeast cells, the three types of sequences required to propagate a chromosome are relatively short (typically less than 1000 base pairs each) and therefore use only a tiny fraction of the information-carrying capacity of a chromosome. Although telomere sequences are fairly simple and short in all eucaryotes, the DNA sequences that specify centromeres and replication origins in more complex organisms are much longer than their yeast counterparts. For example, experiments suggest that human centromeres may contain up to 100,000 nucleotide pairs. It has been proposed that human centromeres may not even require a stretch of DNA with a defined nucleotide sequence; instead, they may simply create a large, regularly repeating protein–nucleic acid structure. We return to this issue at the end of the chapter when we discuss in more general terms the proteins that, along with DNA, make up chromosomes.

DNA Molecules Are Highly Condensed in Chromosomes

All eucaryotic organisms have elaborate ways of packaging DNA into chromosomes. Recall from earlier in this chapter that human chromosome 22 contains about 48 million nucleotide pairs. Stretched out end to end, its DNA would extend about 1.5 cm. Yet, when it exists as a mitotic chromosome, chromosome 22 measures only about 2 μm in length (see Figures 4–10 and 4–11), giving an end-to-end compaction ratio of nearly 10,000-fold. This remarkable feat of compression is performed by proteins that successively coil and fold the DNA into higher and higher levels of organization. Although less condensed than mitotic chromosomes, the DNA of interphase chromosomes is still tightly packed, with an overall compaction ratio of approximately 1000-fold. In the next sections we discuss the specialized proteins that make the compression possible.

In reading these sections it is important to keep in mind that chromosome structure is dynamic. Not only do chromosomes globally condense in accord with the cell cycle, but different regions of the interphase chromosomes condense and decondense as the cells gain access to specific DNA sequences for gene expression, DNA repair, and replication. The packaging of chromosomes must therefore be accomplished in a way that allows rapid localized, on-demand access to the DNA.

Nucleosomes Are the Basic Unit of Eucaryotic Chromosome Structure

The proteins that bind to the DNA to form eucaryotic chromosomes are traditionally divided into two general classes: the **histones** and the *nonhistone chromosomal proteins*. The complex of both classes of protein with the nuclear DNA of eucaryotic cells is known as **chromatin**. Histones are present in such enormous quantities in the cell (about 60 million molecules of each type per human cell) that their total mass in chromatin is about equal to that of the DNA.

Histones are responsible for the first and most basic level of chromosome organization, the **nucleosome**, which was discovered in 1974. When interphase

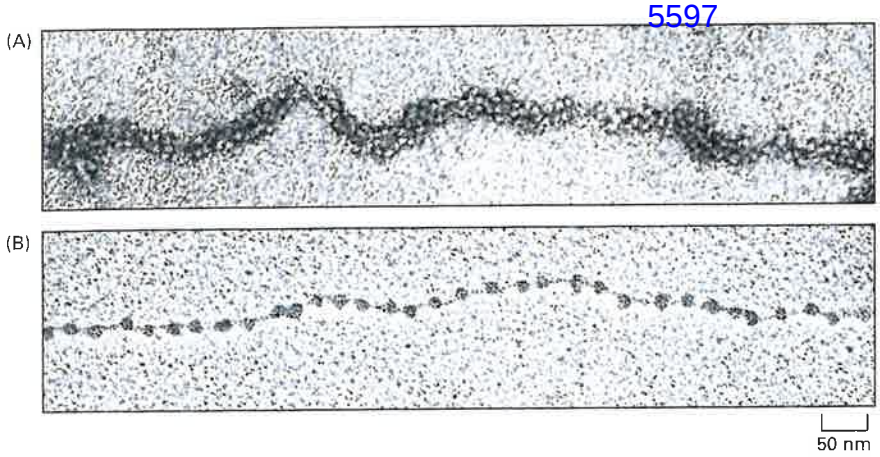


Figure 4–23 Nucleosomes as seen in the electron microscope.
(A) Chromatin isolated directly from an interphase nucleus appears in the electron microscope as a thread 30 nm thick. (B) This electron micrograph shows a length of chromatin that has been experimentally unpacked, or decondensed, after isolation to show the nucleosomes. (A, courtesy of Barbara Hamkalo; B, courtesy of Victoria Foe.)

nuclei are broken open very gently and their contents examined under the electron microscope, most of the chromatin is in the form of a fiber with a diameter of about 30 nm (Figure 4–23A). If this chromatin is subjected to treatments that cause it to unfold partially, it can be seen under the electron microscope as a series of "beads on a string" (Figure 4–23B). The string is DNA, and each bead is a "nucleosome core particle" that consists of DNA wound around a protein core formed from histones. The beads on a string represent the first level of chromosomal DNA packing.

The structural organization of nucleosomes was determined after first isolating them from unfolded chromatin by digestion with particular enzymes (called nucleases) that break down DNA by cutting between the nucleosomes. After digestion for a short period, the exposed DNA between the nucleosome core particles, the *linker DNA*, is degraded. Each individual nucleosome core particle consists of a complex of eight histone proteins—two molecules each of histones H2A, H2B, H3, and H4—and double-stranded DNA that is 146 nucleotide pairs long. The *histone octamer* forms a protein core around which the double-stranded DNA is wound (Figure 4–24).

Each nucleosome core particle is separated from the next by a region of linker DNA, which can vary in length from a few nucleotide pairs up to about 80. (The term *nucleosome* technically refers to a nucleosome core particle plus one of its adjacent DNA linkers, but it is often used synonymously with nucleosome core particle.) On average, therefore, nucleosomes repeat at intervals of about 200 nucleotide pairs. For example, a diploid human cell with 6.4×10^9 nucleotide pairs contains approximately 30 million nucleosomes. The formation of nucleosomes converts a DNA molecule into a chromatin thread about one-third of its initial length, and this provides the first level of DNA packing.

The Structure of the Nucleosome Core Particle Reveals How DNA Is Packaged

The high-resolution structure of a nucleosome core particle, solved in 1997, revealed a disc-shaped histone core around which the DNA was tightly wrapped 1.65 turns in a left-handed coil (Figure 4–25). All four of the histones that make up the core of the nucleosome are relatively small proteins (102–135 amino

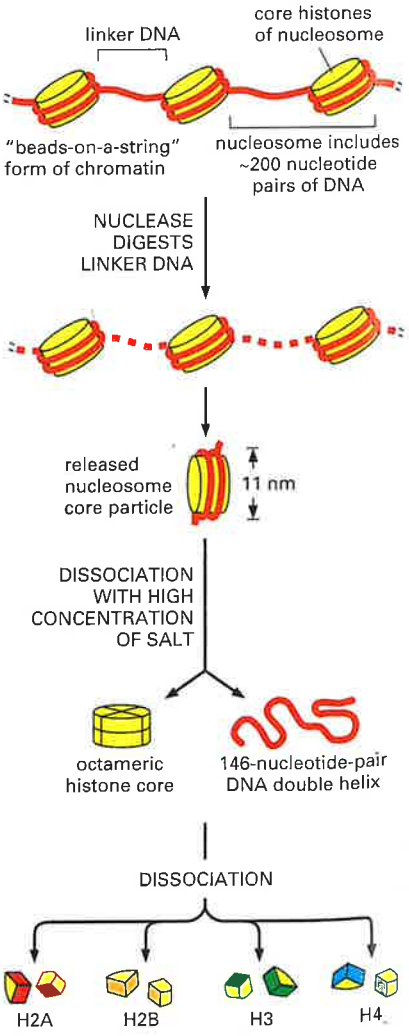


Figure 4–24 Structural organization of the nucleosome.
A nucleosome contains a protein core made of eight histone molecules. As indicated, the nucleosome core particle is released from chromatin by digestion of the linker DNA with a nuclease, an enzyme that breaks down DNA. (The nuclease can degrade the exposed linker DNA but cannot attack the DNA wound tightly around the nucleosome core.) After dissociation of the isolated nucleosome into its protein core and DNA, the length of the DNA that was wound around the core can be determined. This length of 146 nucleotide pairs is sufficient to wrap 1.65 times around the histone core.

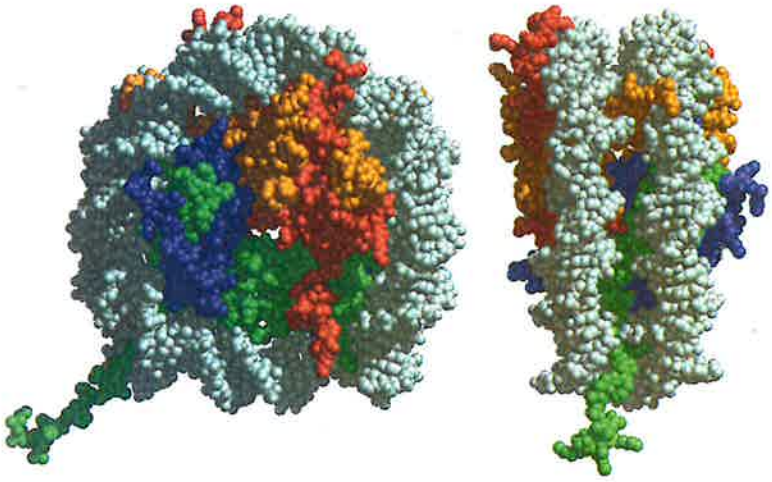


Figure 4–25 The structure of a nucleosome core particle, as determined by x-ray diffraction analyses of crystals. Each histone is colored according to the scheme of Figure 4–24, with the DNA double helix in *light gray*. (Reprinted by permission from K. Luger et al., *Nature* 389:251–260, 1997. © Macmillan Magazines Ltd.)

acids), and they share a structural motif, known as the *histone fold*, formed from three α helices connected by two loops (Figure 4–26). In assembling a nucleosome, the histone folds first bind to each other to form H3–H4 and H2A–H2B dimers, and the H3–H4 dimers combine to form tetramers. An H3–H4 tetramer then further combines with two H2A–H2B dimers to form the compact octamer core, around which the DNA is wound (Figure 4–27).

The interface between DNA and histone is extensive: 142 hydrogen bonds are formed between DNA and the histone core in each nucleosome. Nearly half of these bonds form between the amino acid backbone of the histones and the phosphodiester backbone of the DNA. Numerous hydrophobic interactions and salt linkages also hold DNA and protein together in the nucleosome. For example, all the core histones are rich in lysine and arginine (two amino acids with basic side chains), and their positive charges can effectively neutralize the negatively charged DNA backbone. These numerous interactions explain in part why DNA of virtually any sequence can be bound on a histone octamer core. The path of the DNA around the histone core is not smooth; rather, several kinks are seen in the DNA, as expected from the nonuniform surface of the core.

In addition to its histone fold, each of the core histones has a long N-terminal amino acid "tail", which extends out from the DNA-histone core (see Figure 4–27). These histone tails are subject to several different types of covalent modifications, which control many aspects of chromatin structure. We discuss these issues later in the chapter.

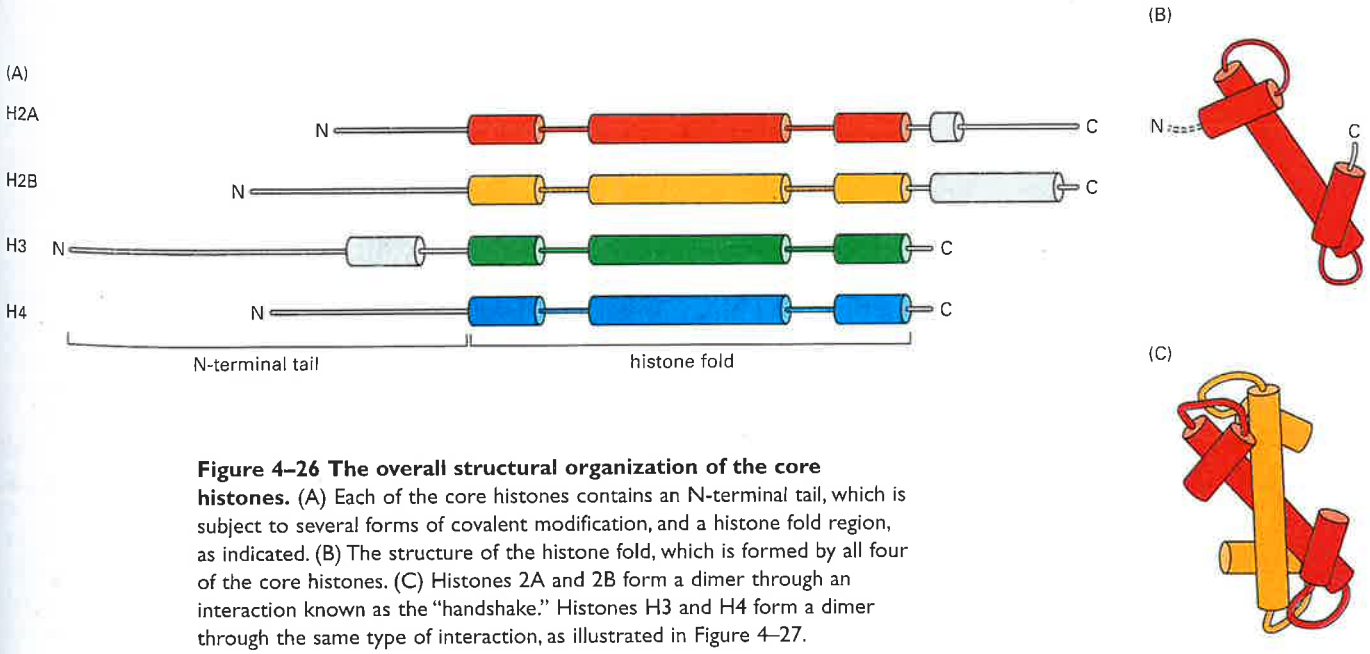


Figure 4–26 The overall structural organization of the core histones. (A) Each of the core histones contains an N-terminal tail, which is subject to several forms of covalent modification, and a histone fold region, as indicated. (B) The structure of the histone fold, which is formed by all four of the core histones. (C) Histones 2A and 2B form a dimer through an interaction known as the "handshake." Histones H3 and H4 form a dimer through the same type of interaction, as illustrated in Figure 4–27.

As might be expected from their fundamental role in DNA packaging, the histones are among the most highly conserved eucaryotic proteins. For example, the amino acid sequence of histone H4 from a pea and a cow differ at only at 2 of the 102 positions. This strong evolutionary conservation suggests that the functions of histones involve nearly all of their amino acids, so that a change in any position is deleterious to the cell. This suggestion has been tested directly in yeast cells, in which it is possible to mutate a given histone gene *in vitro* and introduce it into the yeast genome in place of the normal gene. As might be expected, most changes in histone sequences are lethal; the few that are not lethal cause changes in the normal pattern of gene expression, as well as other abnormalities.

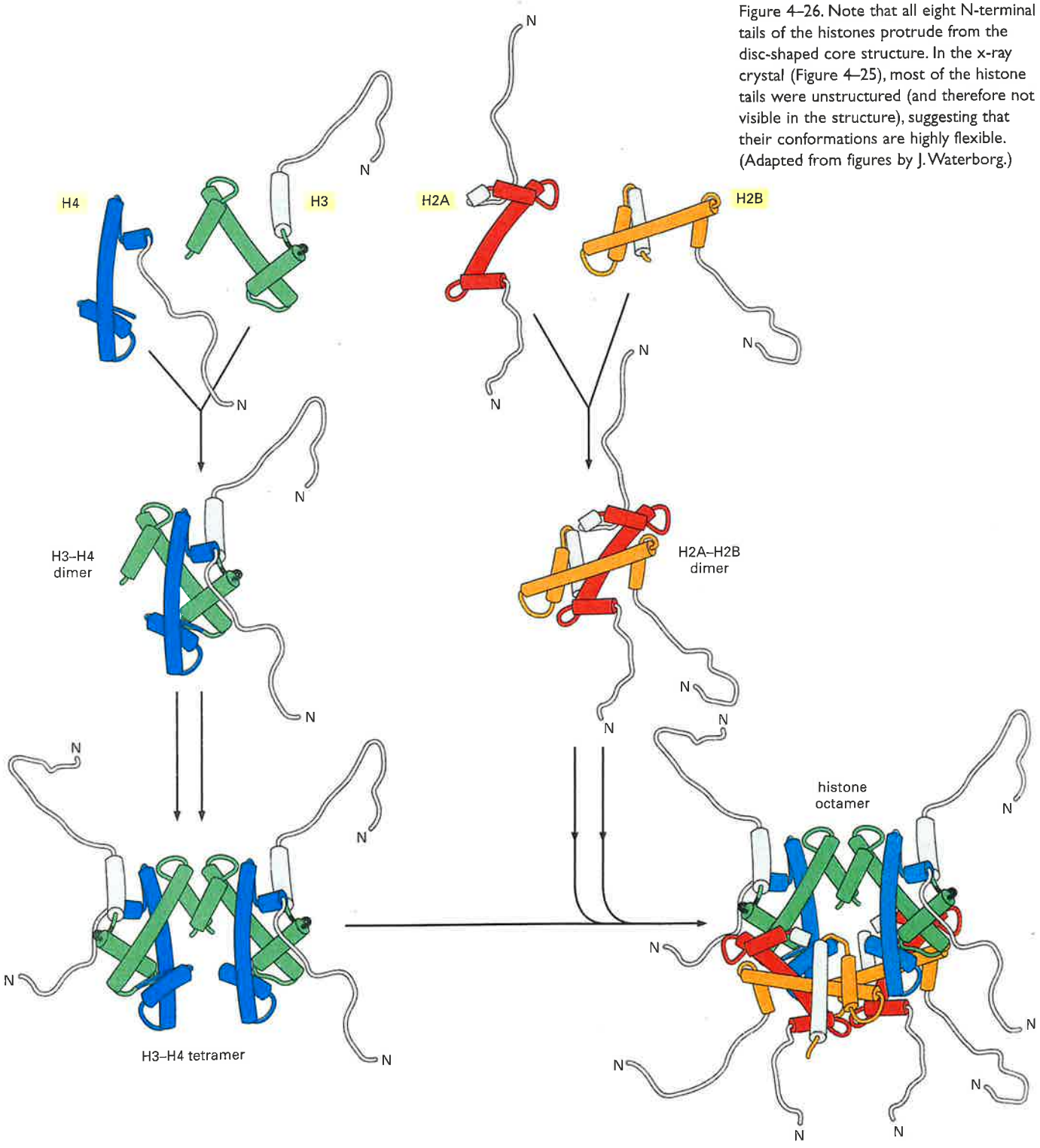


Figure 4–27 The assembly of a histone octamer. The histone H3–H4 dimer and the H2A–H2B dimer are formed from the handshake interaction. An H3–H4 tetramer forms the scaffold of the octamer onto which two H2A–H2B dimers are added, to complete the assembly. The histones are colored as in Figure 4–26. Note that all eight N-terminal tails of the histones protrude from the disc-shaped core structure. In the x-ray crystal (Figure 4–25), most of the histone tails were unstructured (and therefore not visible in the structure), suggesting that their conformations are highly flexible. (Adapted from figures by J. Waterborg.)

Despite the high conservation of the core histones, many eucaryotic organisms also produce specialized variant core histones that differ in amino acid sequence from the main ones. For example, the sea urchin has five histone H2A variants, each of which is expressed at a different time during development. It is thought that nucleosomes that have incorporated these variant histones differ in stability from regular nucleosomes, and they may be particularly well suited for the high rates of DNA transcription and DNA replication that occur during these early stages of development.

The Positioning of Nucleosomes on DNA Is Determined by Both DNA Flexibility and Other DNA-bound Proteins

Although nearly every DNA sequence can, in principle, be folded into a nucleosome, the spacing of nucleosomes in the cell can be irregular. Two main influences determine where nucleosomes form in the DNA. One is the difficulty of bending the DNA double helix into two tight turns around the outside of the histone octamer, a process that requires substantial compression of the minor groove of the DNA helix. Because A-T-rich sequences in the minor groove are easier to compress than G-C-rich sequences, each histone octamer tends to position itself on the DNA so as to maximize A-T-rich minor grooves on the inside of the DNA coil (Figure 4–28). Thus, a segment of DNA that contains short A-T-rich sequences spaced by an integral number of DNA turns is easier to bend around the nucleosome than a segment of DNA lacking this feature. In addition, because the DNA in a nucleosome is kinked in several places, the ability of a given nucleotide sequence to accommodate this deformation can also influence the position of DNA on the nucleosome.

These features of DNA probably explain some striking, but unusual, cases of very precise positioning of nucleosomes along a stretch of DNA. For most of the DNA sequences found in chromosomes, however, there is no strongly preferred nucleosome-binding site; a nucleosome can occupy any one of a number of positions relative to the DNA sequence.

The second, and probably most important, influence on nucleosome positioning is the presence of other tightly bound proteins on the DNA. Some bound proteins favor the formation of a nucleosome adjacent to them. Others create obstacles that force the nucleosomes to assemble at positions between them. Finally, some proteins can bind tightly to DNA even when their DNA-binding site is part of a nucleosome. The exact positions of nucleosomes along a stretch of DNA therefore depend on factors that include the DNA sequence and the presence and nature of other proteins bound to the DNA. Moreover, as we see below, the arrangement of nucleosomes on DNA is highly dynamic, changing rapidly according to the needs of the cell.

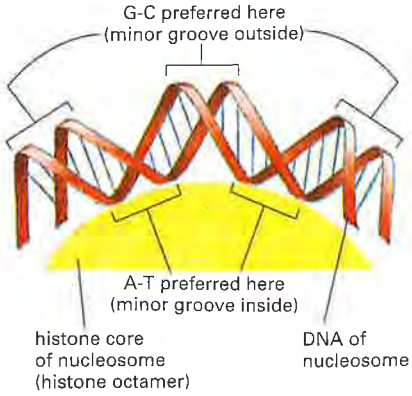


Figure 4–28 The bending of DNA in a nucleosome. The DNA helix makes 1.65 tight turns around the histone octamer. This diagram is drawn approximately to scale, illustrating how the minor groove is compressed on the inside of the turn. Owing to certain structural features of the DNA molecule, A-T base pairs are preferentially accommodated in such a narrow minor groove.

Nucleosomes Are Usually Packed Together into a Compact Chromatin Fiber

Although long strings of nucleosomes form on most chromosomal DNA, chromatin in a living cell probably rarely adopts the extended "beads on a string" form. Instead, the nucleosomes are packed on top of one another, generating regular arrays in which the DNA is even more highly condensed. Thus, when nuclei are very gently lysed onto an electron microscope grid, most of the chromatin is seen to be in the form of a fiber with a diameter of about 30 nm, which is considerably wider than chromatin in the "beads on a string" form (see Figure 4–23).

Several models have been proposed to explain how nucleosomes are packed in the 30-nm chromatin fiber; the one most consistent with the available data is a series of structural variations known collectively as the Zigzag model (Figure 4–29). In reality, the 30-nm structure found in chromosomes is probably a fluid mosaic of the different zigzag variations. We saw earlier that the linker DNA that connects adjacent nucleosomes can vary in length; these differences in linker length probably introduce further local perturbations into the zigzag structure. Finally, the presence of other DNA-binding proteins and DNA sequence that

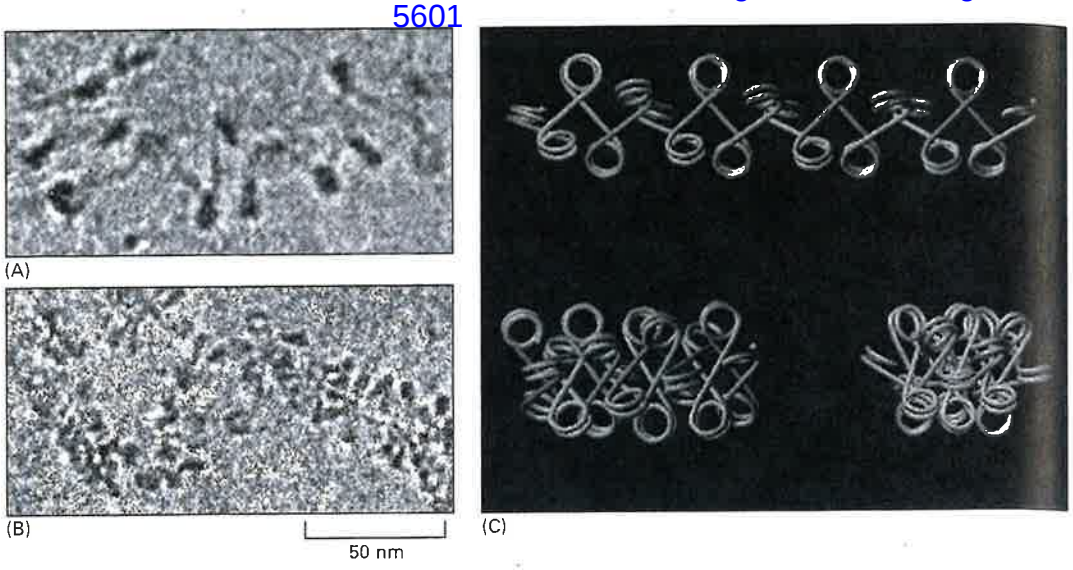


Figure 4–29 Variations on the Zigzag model for the 30-nm chromatin fiber. (A and B) Electron microscopic evidence for the *top* and *bottom-left* model structures depicted in (C). (C) Zigzag variations. An interconversion between these three variations is proposed to occur by an accordion-like expansion and contraction of the fiber length. Differences in the length of the linker between adjacent nucleosome beads can be accommodated by snaking or coiling of the linker DNA, or by small local changes in the width of the fiber. Formation of the 30-nm fiber requires both histone H1 and the core histone tails; for simplicity, neither is shown here, but see Figures 4–30 and 4–32. (From J. Bednar et al., *Proc. Natl. Acad. Sci. USA* 95:14173–14178, 1998. © National Academy of Sciences.)

are difficult to fold into nucleosomes punctuate the 30-nm fiber with irregular features (Figure 4–30).

Several mechanisms probably act together to form the 30-nm fiber from a linear string of nucleosomes. First, an additional histone, called histone H1, is involved in this process. H1 is larger than the core histones and is considerably less well conserved. In fact, the cells of most eucaryotic organisms make several histone H1 proteins of related but quite distinct amino acid sequences. A single histone H1 molecule binds to each nucleosome, contacting both DNA and protein, and changing the path of the DNA as it exits from the nucleosome. Although it is not understood in detail how H1 pulls nucleosomes together into the 30-nm fiber, a change in the exit path in DNA seems crucial for compacting nucleosomal DNA so that it interlocks to form the 30-nm fiber (Figure 4–31).

A second mechanism for forming the 30-nm fiber probably involves the tails of the core histones, which, as we saw above, extend from the nucleosome. It is thought that these tails may help attach one nucleosome to another—thereby allowing a string of them, with the aid of histone H1, to condense into the 30-nm fiber (Figure 4–32).

ATP-driven Chromatin Remodeling Machines Change Nucleosome Structure

For many years biologists thought that, once formed in a particular position on DNA, a nucleosome remained fixed in place because of the tight association

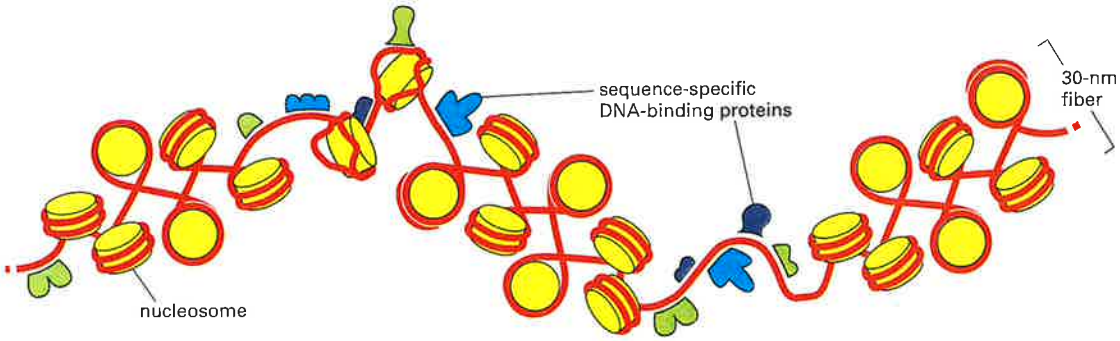


Figure 4–30 Irregularities in the 30-nm fiber. This schematic view of the 30-nm fiber illustrates its interruption by sequence-specific DNA-binding proteins. How these proteins bind tightly to DNA is explained in Chapter 7. The interruptions in the 30-nm fiber may be due to regions of DNA that lack nucleosomes altogether or, more probably, to regions that contain altered or remodeled nucleosomes. Regions of chromatin that are nucleosome free or contain remodeled nucleosome can often be detected experimentally by the unusually high susceptibility of their DNA to digestion by nucleases—as compared with the DNA in nucleosomes.

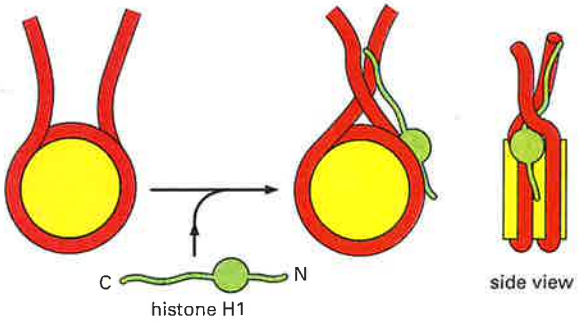


Figure 4–31 A speculative model for how histone H1 could change the path of DNA as it exits from the nucleosome. Histone H1 *(green)* consists of a globular core and two extended tails. Part of the effect of H1 on the compaction of nucleosome organization may result from charge neutralization: like the core histones, H1 is positively charged (especially its C-terminal tail), and this helps to compact the negatively charged DNA. Unlike the core histones, H1 does not seem to be essential for cell viability; in one ciliated protozoan the nucleus expands nearly twofold in the absence of H1, but the cells otherwise appear normal.

between the core histones and DNA. But it has recently been discovered that eucaryotic cells contain *chromatin remodeling complexes*, protein machines that use the energy of ATP hydrolysis to change the structure of nucleosomes temporarily so that DNA becomes less tightly bound to the histone core. The remodeled state may result from movement of the H2A–H2B dimers in the nucleosome core; the H3–H4 tetramer is particularly stable and would be difficult to rearrange (see Figure 4–27).

The remodeling of nucleosome structure has two important consequences. First, it permits ready access to nucleosomal DNA by other proteins in the cell, particularly those involved in gene expression, DNA replication, and repair. Even after the remodeling complex has dissociated, the nucleosome can remain in a "remodeled state" that contains DNA and the full complement of histones—but one in which the DNA–histone contacts have been loosened; only gradually does this remodeled state revert to that of a standard nucleosome. Second, remodeling complexes can catalyze changes in the positions of nucleosomes along DNA (Figure 4–33); some can even transfer a histone core from one DNA molecule to another.

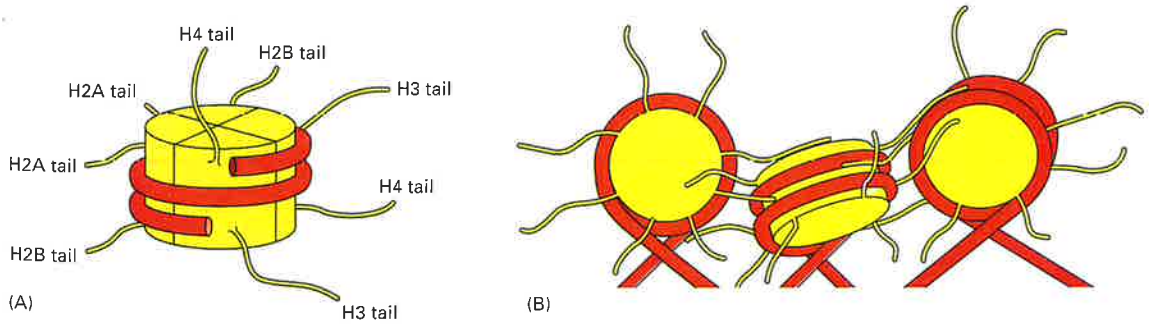
Cells have several different chromatin remodeling complexes that differ subtly in their properties. Most are large protein complexes that can contain more than ten subunits. It is likely that they are used whenever a eucaryotic cell needs direct access to nucleosome DNA for gene expression, DNA replication, or DNA repair. Different remodeling complexes may have features specialized for each of these roles. It is thought that the primary role of some remodeling complexes is to allow access to nucleosomal DNA, whereas that of others is to re-form nucleosomes when access to DNA is no longer required (Figure 4–34).

Chromatin remodeling complexes are carefully controlled by the cell. We shall see in Chapter 7 that, when genes are turned on and off, these complexes can be brought to specific regions of DNA where they act locally to influence chromatin structure. During mitosis, at least some of the chromatin-remodeling complexes are inactivated by phosphorylation. This may help the tightly packaged mitotic chromosomes maintain their structure.

Covalent Modification of the Histone Tails Can Profoundly Affect Chromatin

The N-terminal tails of each of the four core histones are highly conserved in their sequence, and perform crucial functions in regulating chromatin structure.

Figure 4–32 A speculative model for histone tails in the formation of the 30-nm fiber. (A) The approximate exit points of the eight histone tails, four from each histone subunit, that extend from each nucleosome. In the high-resolution structure of the nucleosome (see Figure 4–25), the tails are largely unstructured, suggesting that they are highly flexible. (B) A speculative model showing how the histone tails may help to pack nucleosomes together into the 30-nm fiber. This model is based on (1) experimental evidence that histone tails aid in the formation of the 30-nm fiber, (2) the x-ray crystal structure of the nucleosome, which showed that the tails of one nucleosome contact the histone core of an adjacent nucleosome in the crystal lattice, and (3) evidence that the histone tails interact with DNA.



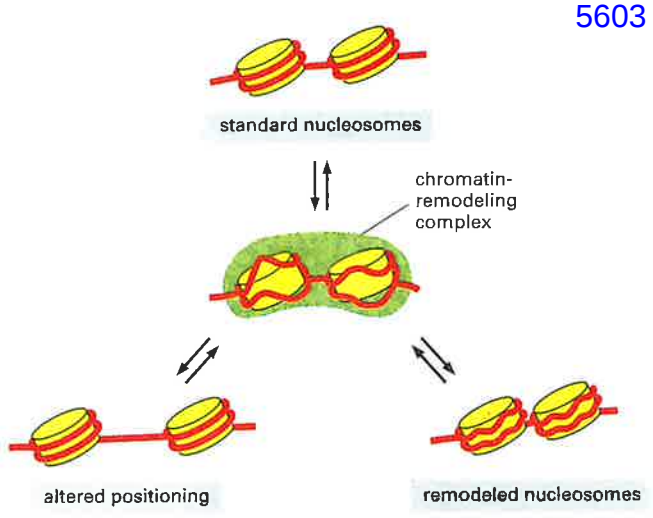


Figure 4–33 Model for the mechanism of some chromatin remodeling complexes. In the absence of remodeling complexes, the interconversion between the three nucleosomal states shown is very slow because of a high activation energy barrier. Using ATP hydrolysis, chromatin-remodeling complexes *(green)* create an activated intermediate (shown in the center of the figure) in which the histone–DNA contacts have been partly disrupted. This activated state can then decay to any one of the three nucleosomal configurations shown. In this way, the remodeling complexes greatly increase the rate of interconversion between different nucleosomal states. The remodeled state, in which the histone–DNA contacts have been loosened, has a higher free energy level than that of standard nucleosomes and will slowly revert to the standard nucleosome conformation, even in the absence of a remodeling complex. Cells have many different chromatin remodeling complexes, and they differ in their detailed biochemical properties; for example, not all can change the position of a nucleosome, but all use the energy of ATP hydrolysis to alter nucleosome structure. (Adapted from R.E. Kingston and G.J. Narlikar, *Genes Dev.* 13:2339–2352, 1999.)

Each tail is subject to several types of covalent modifications, including acetylation of lysines, methylation of lysines, and phosphorylation of serines (Figure 4–35A). Histones are synthesized in the cytosol and then assembled into nucleosomes. Some of the modifications of histone tails occur just after their synthesis, but before their assembly. The modifications that concern us, however, take place once the nucleosome has been assembled. These nucleosome modifications are added and removed by enzymes that reside in the nucleus; for example, acetyl groups are added to the histone tails by histone acetyl transferases (HATs) and taken off by histone deacetylases (HDACs).

The various modifications of the histone tails have several important consequences. Although modifications of the tails have little direct effect on the stability of an individual nucleosome, they seem to affect the stability of the

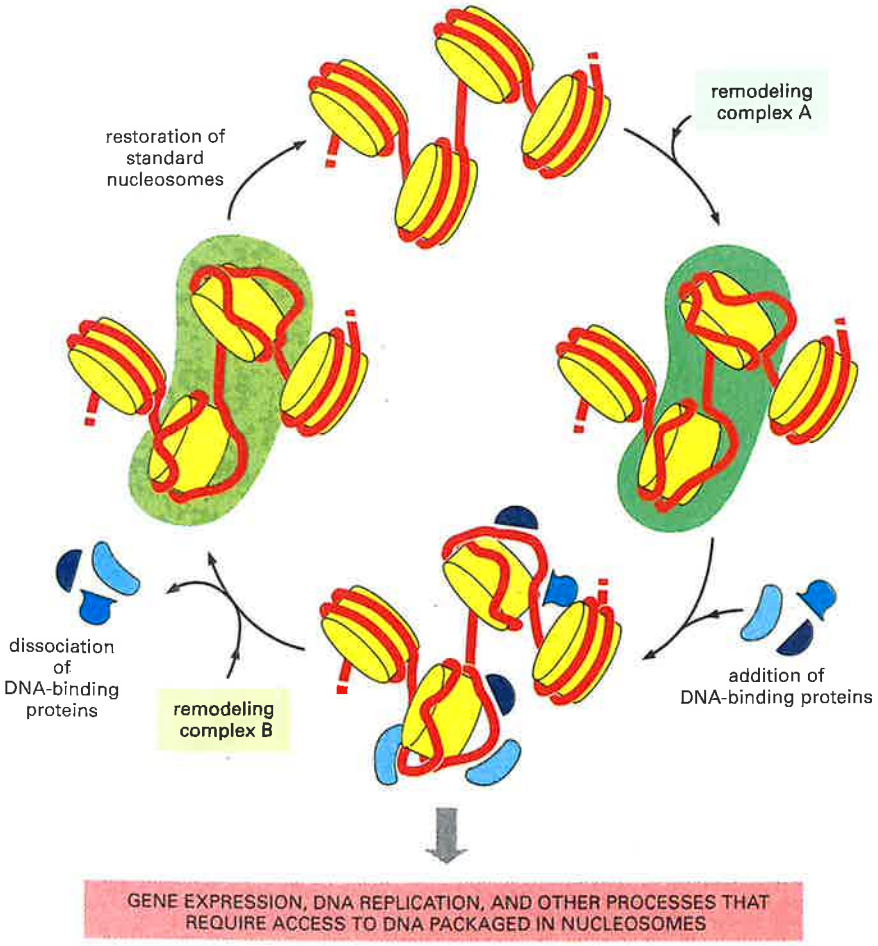
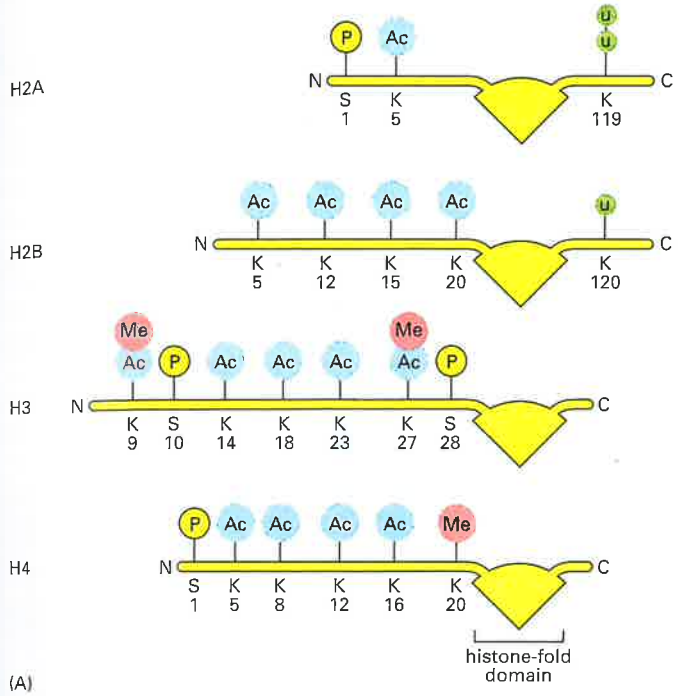
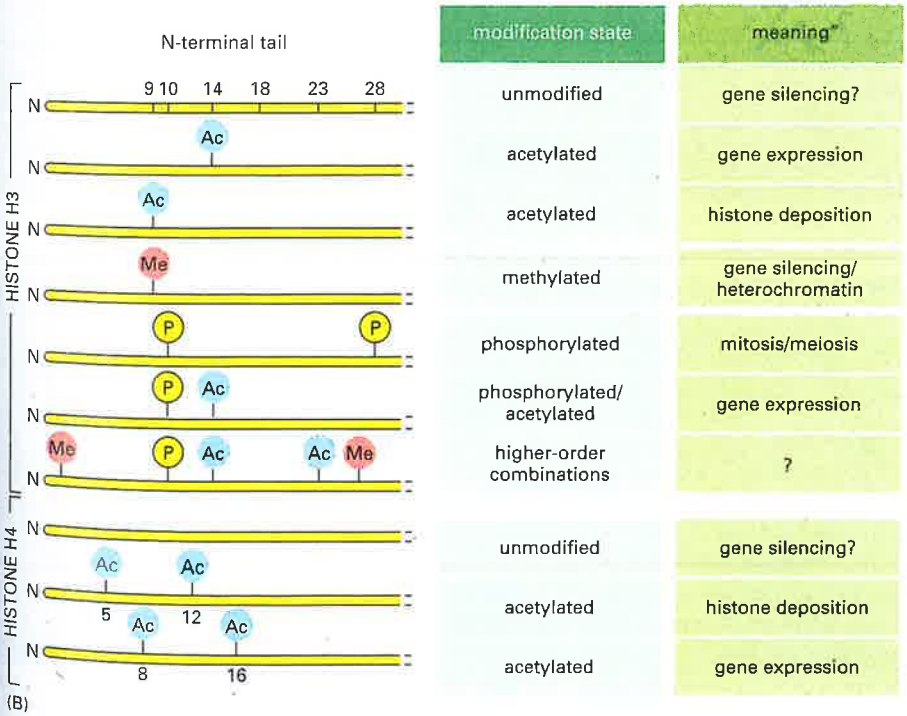


Figure 4–34 A cyclic mechanism for nucleosome disruption and re-formation. According to this model, different chromatin remodeling complexes disrupt and re-form nucleosomes, although, in principle, the same complex might catalyze both reactions. The DNA-binding proteins could function in gene expression, DNA replication, or DNA repair, and in some cases their binding could lead to the dissociation of the histone core to form nucleosome-free regions of DNA like those illustrated in Figure 4–30. (Adapted from A. Travers, *Cell* 96:311–314, 1999.)

30-nm chromatin fiber and of the higher-order structures discussed below. For example, histone acetylation tends to destabilize chromatin structure, perhaps in part because adding an acetyl group removes the positive charge from the lysine, thereby making it more difficult for histones to neutralize the charges on DNA as chromatin is compacted. However, the most profound effect of modified histone tails is their ability to attract specific proteins to a stretch of chromatin that has been appropriately modified. Depending on the precise tail modifications, these additional proteins can either cause further compaction of the chromatin or can facilitate access to the DNA. If combinations of modifications are taken into account, the number of possible distinct markings for each histone tail is very large. Thus, it has been proposed that, through covalent modification of the histone tails, a given stretch of chromatin can convey a particular meaning to





modification state	meaning*
unmodified	gene silencing?
acetylated	gene expression
acetylated	histone deposition
methylated	gene silencing/ heterochromatin
phosphorylated	mitosis/meiosis
phosphorylated/ acetylated	gene expression
higher-order combinations	?
unmodified	gene silencing?
acetylated	histone deposition
acetylated	gene expression

Figure 4–35 Covalent modification of core histone tails. (A) Known modifications of the four histone core proteins are indicated: Me = methyl group, Ac = acetyl group, P = phosphate, u = ubiquitin. Note that some positions (e.g., lysine 9 of H3) can be modified in more than one way. Most of these modifications add a relatively small molecule onto the histone tails; the exception is ubiquitin, a 76 amino acid protein also used in other cellular processes (see Figure 6–87). The function of ubiquitin in chromatin is not well understood: histone H2B can be modified by a single ubiquitin molecule; H2A can be modified by the addition of several ubiquitins. (B) A histone code hypothesis. Histone tails can be marked by different combinations of modifications. According to this hypothesis, each marking conveys a specific meaning to the stretch of chromatin on which it occurs. Only a few of the meanings of the modifications are known. In Chapter 7, we discuss the way a doubly-acetylated H4 tail is "read" by a protein required for gene expression. In another well-studied case, an H3 tail methylated at lysine 9 is recognized by a set of proteins that create an especially compact form of chromatin, which silences gene expression.

The acetylation of lysine 14 of histone H3 and lysines 8 and 16 of histone H4—usually associated with gene expression—is performed by the type A histone acetylases (HATs) in the nucleus. In contrast, the acetylation of lysines 5 and 12 of histone H4 and a lysine of histone H3 takes place in the cytosol, after the histones have been synthesized but before they have been incorporated into nucleosomes; these modifications are catalyzed by type B HATs. These modified histones are deposited onto DNA after DNA replication (see Figure 5–41), and their acetyl groups are taken off shortly afterwards by histone deacetylases (HDACs). Thus, the acetylation at these positions signals newly replicated chromatin.

Modification of a particular position in a histone tail can take on different meanings depending on other features of the local chromatin structure. For example, the phosphorylation of position 10 of histone H3 is associated not only with the condensation of chromosomes that takes place in mitosis and meiosis but also with the expression of certain genes. Some histone tail modifications are interdependent. For example methylation of H3 position 9 blocks the phosphorylation of H3 position 10, and vice versa.

the cell (Figure 4–35B). For example, one type of marking could signal that the stretch of chromatin has been newly replicated, and another could signal that gene expression should not take place. According to this idea, each different marking would attract those proteins that would then execute the appropriate functions. Because the histone tails are extended, and are therefore probably accessible even when chromatin is condensed, they provide an especially apt format for such messages.

As with chromatin remodeling complexes, the enzymes that modify (and remove modifications from) histone tails are usually multisubunit proteins, and they are tightly regulated. They are brought to a particular region of chromatin by other cues, particularly by sequence-specific DNA-binding proteins. We can thus imagine how cycles of histone tail modification and demodification can allow chromatin structure to be dynamic—locally compacting and decompacting it, and, in addition, attracting other proteins specific for each modification state. It is likely that histone-modifying enzymes and chromatin remodeling complexes work in concert to condense and recondense stretches of chromatin; for example, evidence suggests that a particular modification of the histone tail attracts a particular type of remodeling complex. Moreover, some chromatin remodeling complexes contain histone modification enzymes as subunits, directly connecting the two processes.

Summary

A gene is a nucleotide sequence in a DNA molecule that acts as a functional unit for the production of a protein, a structural RNA, or a catalytic RNA molecule. In eucaryotes, protein-coding genes are usually composed of a string of alternating introns and exons. A chromosome is formed from a single, enormously long DNA molecule that contains a linear array of many genes. The human genome contains 3.2×10^9 DNA nucleotide pairs, divided between 22 different autosomes and 2 sex chromosomes. Only a small percentage of this DNA codes for proteins or structural and catalytic RNAs. A chromosomal DNA molecule also contains three other types of functionally important nucleotide sequences: replication origins and telomeres allow the DNA molecule to be completely replicated, while a centromere attaches the daughter DNA molecules to the mitotic spindle, ensuring their accurate segregation to daughter cells during the M phase of the cell cycle.

The DNA in eucaryotes is tightly bound to an equal mass of histones, which form a repeating array of DNA–protein particles called nucleosomes. The nucleosome is composed of an octameric core of histone proteins around which the DNA double helix is wrapped. Despite irregularities in the positioning of nucleosomes along DNA, nucleosomes are usually packed together (with the aid of histone H1 molecules) into quasi-regular arrays to form a 30-nm fiber. Despite the high degree of compaction in chromatin, its structure must be highly dynamic to allow the cell access to the DNA. Two general strategies for reversibly changing local chromatin structures are important for this purpose: ATP-driven chromatin remodeling complexes, and an enzymatically catalyzed covalent modification of the N-terminal tails of the four core histones.

THE GLOBAL STRUCTURE OF CHROMOSOMES

Having discussed the DNA and protein molecules from which the 30-nm chromatin fiber is made, we now turn to the organization of the chromosome on a more global scale. As a 30-nm fiber, the typical human chromosome would still be 0.1 cm in length and able to span the nucleus more than 100 times. Clearly, there must be a still higher level of folding, even in interphase chromosomes. This higher-order packaging is one of the most fascinating—but also one of the most poorly understood—aspects of chromosome structure. Although its molecular basis is still largely a mystery, it almost certainly involves the folding of the 30-nm fiber into a series of loops and coils, as we see below. Our discussion of this higher-order packing continues an important theme in chromosome

architecture: interphase chromatin structure is fluid, exposing at any given moment the DNA sequences directly needed by the cell.

We first describe several rare cases in which the overall structure and organization of interphase chromosomes can be easily visualized, and we explain that certain features of these exceptional cases may be representative of the structures of all interphase chromosomes. Next we describe the different forms of chromatin that make up a typical interphase chromosome. Finally we discuss the additional compaction that interphase chromosomes undergo during the process of mitosis.

Lampbrush Chromosomes Contain Loops of Decondensed Chromatin

Most chromosomes in interphase cells are too fine and too tangled to be visualized clearly. In a few exceptional cases, however, *interphase chromosomes* can be seen to have a precisely defined higher-order structure, and it is thought that certain characteristics of these higher-order structures are representative of all interphase chromosomes. The meiotically paired chromosomes in growing amphibian oocytes (immature eggs), for example, are highly active in gene expression, and they form unusually stiff and extended chromatin loops. These so-called **lampbrush chromosomes** (the largest chromosomes known) are clearly visible even in the light microscope, where they are seen to be organized into a series of large chromatin loops emanating from a linear chromosomal axis (Figure 4–36).

The organization of a lampbrush chromosome is illustrated in Figure 4–37. A given loop always contains the same DNA sequence, and it remains extended in the same manner as the oocyte grows. Other experiments demonstrate that

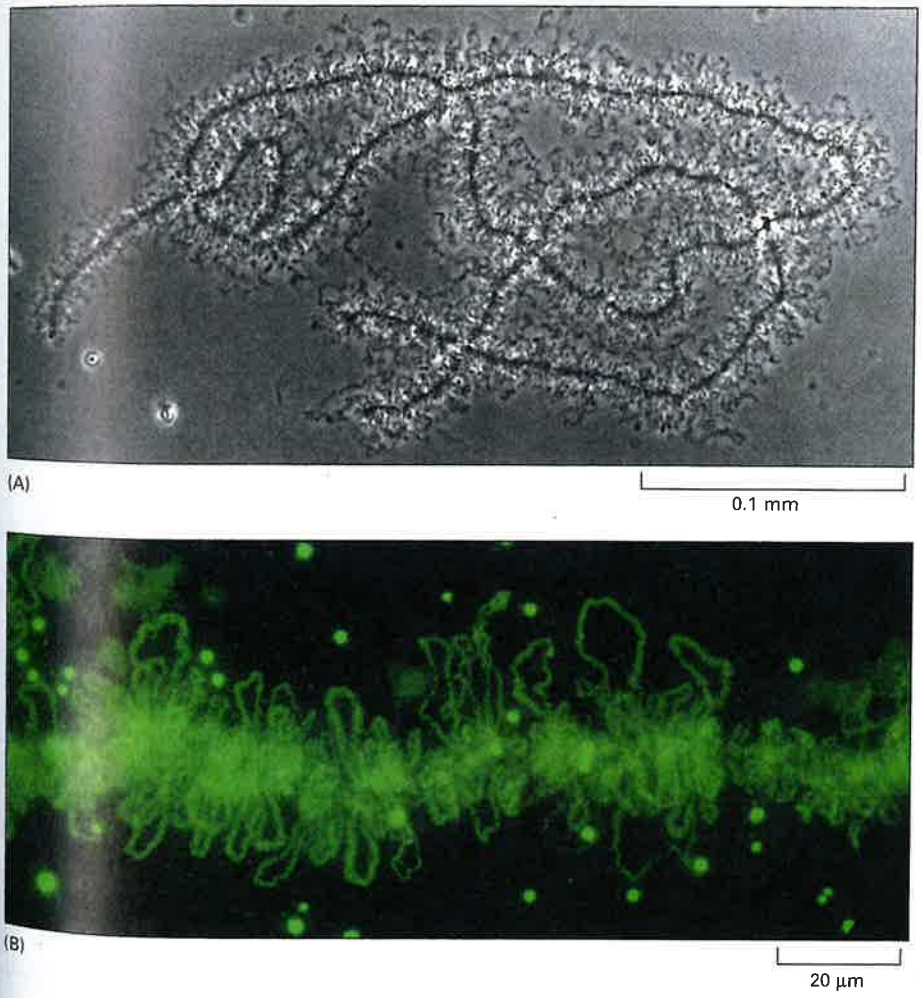


Figure 4–36 Lampbrush chromosomes. (A) A light micrograph of lampbrush chromosomes in an amphibian oocyte. Early in oocyte differentiation, each chromosome replicates to begin meiosis, and the homologous replicated chromosomes pair to form this highly extended structure containing a total of four replicated DNA molecules, or chromatids. The lampbrush chromosome stage persists for months or years, while the oocyte builds up a supply of materials required for its ultimate development into a new individual. (B) Fluorescence light micrograph showing a portion of an amphibian lampbrush chromosome. The regions of the chromosome that are being actively expressed are stained *green* by using antibodies against proteins that process RNA during one of the steps of gene expression (discussed in Chapter 6). The round granules are thought to correspond to large complexes of the RNA-splicing machinery that will also be discussed in Chapter 6. (A, courtesy of Joseph G. Gall; B, courtesy of Joseph G. Gall and Christine Murphy.)

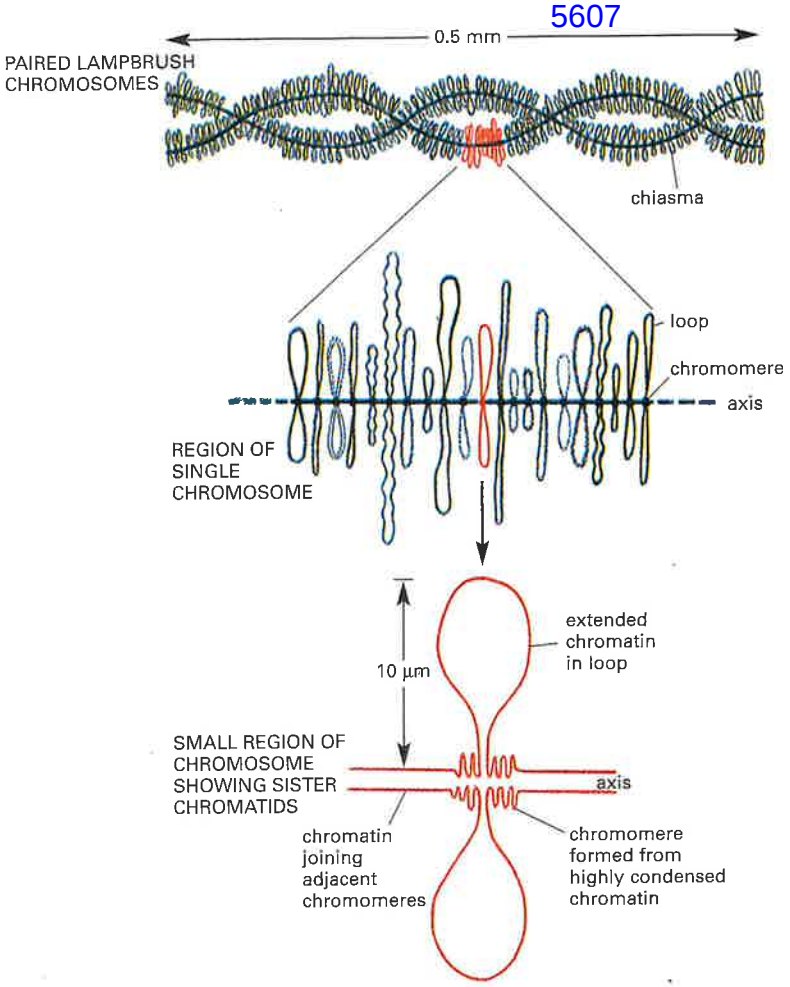


Figure 4–37 A model for the structure of a lampbrush chromosome. The set of lampbrush chromosomes in many amphibians contains a total of about 10,000 chromatin loops, although most of the DNA in each chromosome remains highly condensed in the chromomeres. Each loop corresponds to a particular DNA sequence. Four copies of each loop are present in each cell, since each of the two chromosomes shown at the top consists of two closely apposed, newly replicated chromosomes. This four-stranded structure is characteristic of this stage of development of the oocyte, the diplotene stage of meiosis; see Figure 20–12.

most of the genes present in the DNA loops are being actively expressed (see Figure 4–36B). Most of the DNA, however, is not in loops but remains highly condensed in the *chromomeres* on the axis, which are generally not expressed. Lampbrush chromosomes illustrate a recurrent theme of this chapter—when the DNA in a region of chromatin is in use (in this case, for gene expression), that part of the chromatin has an extended structure; otherwise, the chromatin is condensed. In lampbrush chromosomes, the structural units of this regulation are large, precisely defined loops.

Relatively few species undergo the specialization that produces lampbrush chromosomes. However, when injected into amphibian oocytes, the DNA from organisms that normally do not produce lampbrush chromosomes (e.g., DNA from a fish) is packaged into lampbrush chromosomes. On the basis of this type of experiment, it has been proposed that the interphase chromosomes of all eucaryotes are arranged in loops that are normally too small and fragile to be easily observed. It may be possible in the future to coax the DNA from a mammal such as a mouse to form lampbrush chromosomes by introducing it into amphibian oocytes. This could allow a detailed correlation of loop structure, gene arrangement, and DNA sequence, and we could begin to learn how the packaging into loops reflects the sequence content of our DNA.

Drosophila Polytene Chromosomes Are Arranged in Alternating Bands and Interbands

Certain insect cells also have specialized interphase chromosomes that are readily visible, although this type of specialization differs from that of lampbrush chromosomes. For example, many of the cells of certain fly larvae grow to an enormous size through multiple cycles of DNA synthesis without cell division. The resulting giant cells contain as much as several thousand times the normal DNA complement. Cells with more than the normal DNA complement are said

to be *polyploid* when they contain increased numbers of standard chromosomes. In several types of secretory cells of fly larvae, however, all the homologous chromosome copies are held side by side, like drinking straws in a box, creating a single **polytene chromosome**. The fact that, in some large insect cells, polytene chromosomes can disperse to form a conventional polyploid cell demonstrates that these two chromosomal states are closely related, and that the basic structure of a polytene chromosome must be similar to that of a normal chromosome.

Polytene chromosomes are often easy to see in the light microscope because they are so large and because the precisely aligned side-by-side adherence of individual chromatin strands greatly elongates the chromosome axis and prevents tangling. Polyteny has been most studied in the salivary gland cells of *Drosophila* larvae, in which the DNA in each of the four *Drosophila* chromosomes has been replicated through 10 cycles without separation of the daughter chromosomes, so that 1024 (2^{10}) identical strands of chromatin are lined up side by side (Figure 4–38).

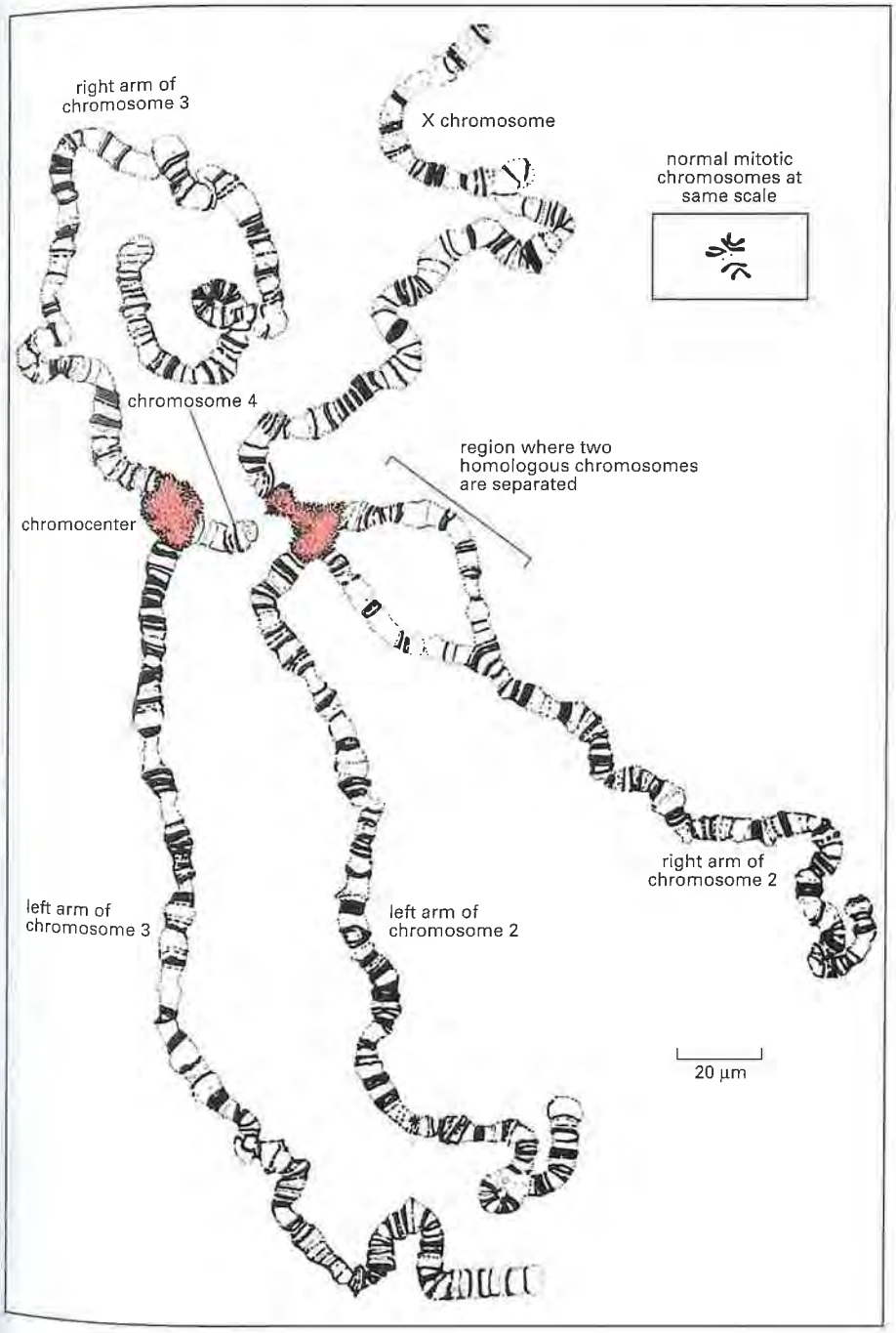


Figure 4–38 The entire set of polytene chromosomes in one *Drosophila* salivary cell. These chromosomes have been spread out for viewing by squashing them against a microscope slide. *Drosophila* has four chromosomes, and there are four different chromosome pairs present. But each chromosome is tightly paired with its homolog (so that each pair appears as a single structure), which is not true in most nuclei (except in meiosis). The four polytene chromosomes are normally linked together by regions near their centromeres that aggregate to create a single large chromocenter *(pink region)*. In this preparation, however, the chromocenter has been split into two halves by the squashing procedure used. (Modified from T.S. Painter, *J. Hered.* 25:465–476, 1934.)

When polytene chromosomes are viewed in the light microscope, distinct alternating dark *bands* and light *interbands* are visible (Figure 4–39). Each band and interband represents a set of 1024 identical DNA sequences arranged in register. About 95% of the DNA in polytene chromosomes is in bands, and 5% is in interbands. The chromatin in each band appears dark, either because it is much more condensed than the chromatin in the interbands, or because it contains a higher proportion of proteins, or both (Figure 4–40). Depending on their size, individual bands are estimated to contain 3000–300,000 nucleotide pairs in a chromatin strand. The bands of *Drosophila* polytene chromosomes can be recognized by their different thicknesses and spacings, and each one has been given a number to generate a chromosome "map." There are approximately 5000 bands and 5000 interbands in the complete set of *Drosophila* polytene chromosomes.

Both Bands and Interbands in Polytene Chromosomes Contain Genes

The reproducible pattern of bands and interbands seen in *Drosophila* polytene chromosomes means that these interphase chromosomes are highly organized. Since the 1930s, scientists have debated the nature of this organization, and we still do not have a clear answer. Because the number of bands in *Drosophila* chromosomes was once thought to be roughly equal to the number of genes in the genome, it was initially thought that each band might correspond to a single gene; however, we now know this simple idea is incorrect. There are nearly three times more genes in *Drosophila* than chromosome bands, and genes are found in both band and interband regions. Moreover, some bands contain multiple genes, and some bands seem to lack genes altogether.

It seems likely that the band–interband pattern reflects different levels of gene expression and chromatin structure along the chromosome, with genes in the less compact interbands being expressed more highly than those in the more compact bands. In any case, the remarkable appearance of fly polytene chromosomes is thought to reflect the heterogeneous nature of the chromatin compaction found along all interphase chromosomes. In the next section we see how the appearance of a band can change dramatically when the gene or genes within it become highly expressed.

Individual Polytene Chromosome Bands Can Unfold and Refold as a Unit

A major factor controlling gene expression in the polytene chromosomes of *Drosophila* is the insect steroid hormone *ecdysone*, the levels of which rise and fall periodically during larval development. When ecdysone concentrations rise,

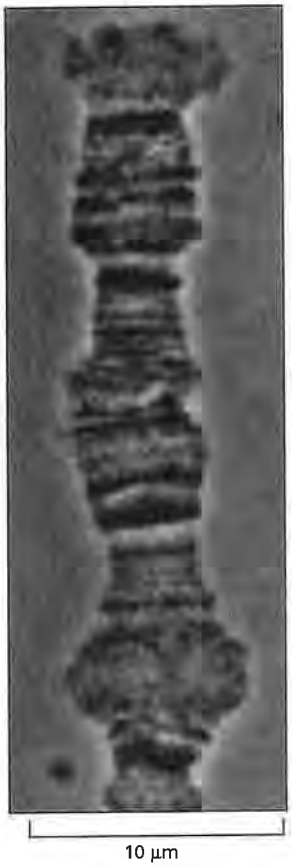


Figure 4–39 A light micrograph of a portion of a polytene chromosome from *Drosophila* salivary glands. The distinct pattern produced by bands and interbands is readily seen. The bands are regions of increased chromatin concentration that occur in interphase chromosomes. Although they are detectable only in polytene chromosomes, it is thought that they reflect a structure common to the chromosomes of most eucaryotes. (Courtesy of Joseph G. Gall.)

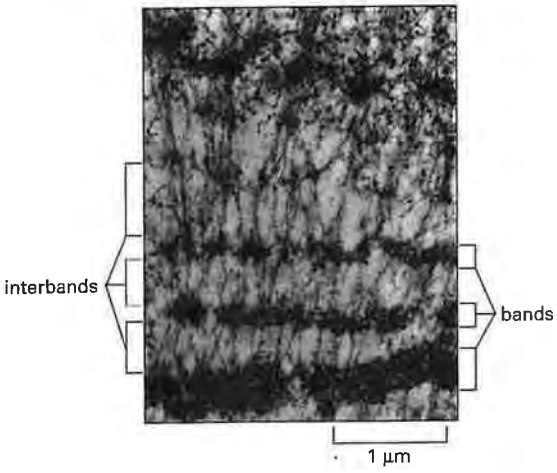


Figure 4–40 An electron micrograph of a small section of a *Drosophila* polytene chromosome seen in thin section. Bands of very different thickness can be readily distinguished, separated by interbands, which contain less condensed chromatin. (Courtesy of Veikko Sorsa.)

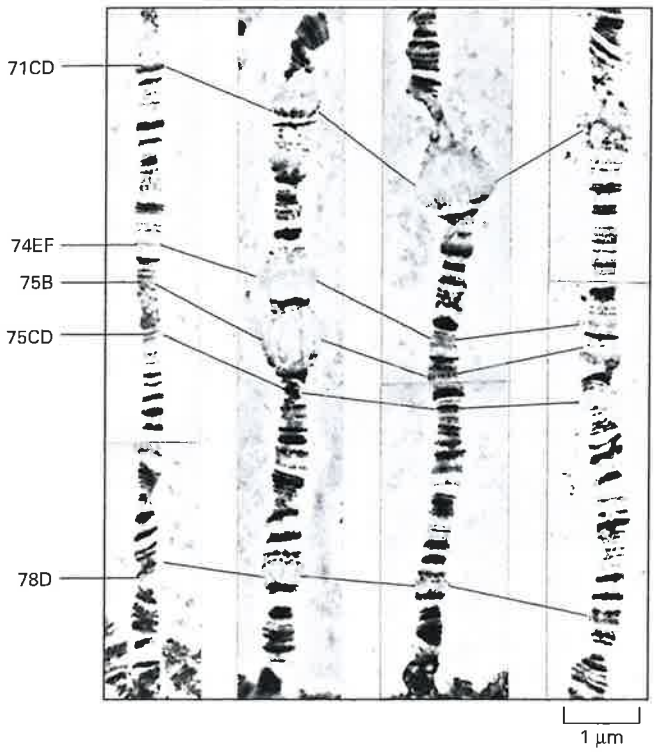


Figure 4–41 Chromosome puffs. This series of time-lapse photographs shows how puffs arise and recede in the polytene chromosomes of *Drosophila* during larval development. A region of the *left* arm of chromosome 3 is shown. It exhibits five very large puffs in salivary gland cells, each active for only a short developmental period. The series of changes shown occur over a period of 22 hours, appearing in a reproducible pattern as the organism develops. The designations of the indicated bands are given at the left of the photographs. (Courtesy of Michael Ashburner.)

they induce the expression of genes coding for proteins that the larva requires for each molt and for pupation. As the organism progresses from one developmental stage to another, distinctive *chromosome puffs* arise and old puffs recede as new genes become expressed and old ones are turned off (Figure 4–41). From inspection of each puff when it is relatively small and the banding pattern is still discernible, it seems that most puffs arise from the decondensation of a single chromosome band.

The individual chromatin fibers that make up a puff can be visualized with an electron microscope. For technical reasons, this is easier in the polytene chromosomes from a different insect, *Chironomus tentans*, a midge. Electron micrographs of certain puffs, called Balbiani rings, of *Chironomus* salivary gland polytene chromosomes show the chromatin arranged in loops (Figures 4–42 and 4–43), much like those observed in the amphibian lampbrush chromosomes discussed earlier. Additional experiments suggest that each loop contains a

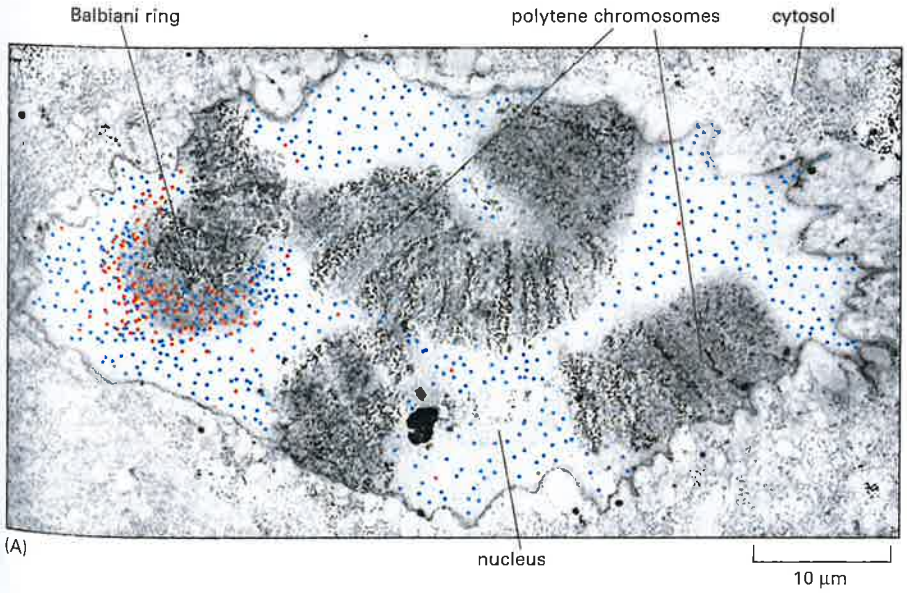
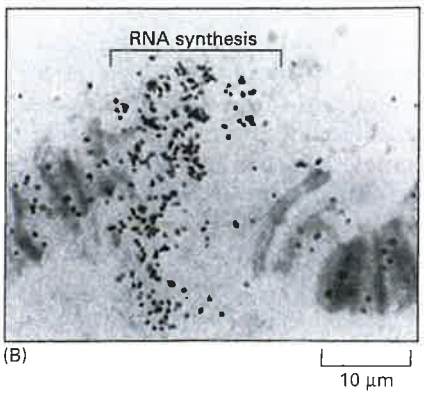


Figure 4–42 RNA synthesis in chromosome puffs. (A) Polytene chromosomes from the salivary glands of the insect *C. tentans*. As outlined in Chapter 1 and described in detail in Chapter 6, the first step in gene expression is the synthesis of an RNA molecule using the DNA as a template. In this electron micrograph, newly synthesized RNA from a Balbiani ring gene is indicated in *red*. Cells were exposed to a brief pulse of BrUTP (an analog of UTP), which was incorporated into RNA. Cells were then fixed and the newly synthesized RNAs were identified by using antibodies against BrU. Balbiani ring RNAs could be distinguished from other RNAs by their characteristic shape (see Figure 4–43). The *blue dots* in the figure represent positions of Balbiani ring RNAs that were synthesized before the addition of BrUTP. The experiment shows that Balbiani ring RNAs are synthesized in puffs and then diffuse through the nucleoplasm. (B) An autoradiogram of a single puff in a polytene chromosome. The portion of the chromosome indicated is undergoing RNA synthesis and has therefore become labeled with ^{3}H-uridine. (A, courtesy of B. Daneholt, from O.P Singh et al., *Exp. Cell Res.* 251:135–146, 1999. © Academic Press; B, courtesy of José Bonner.)



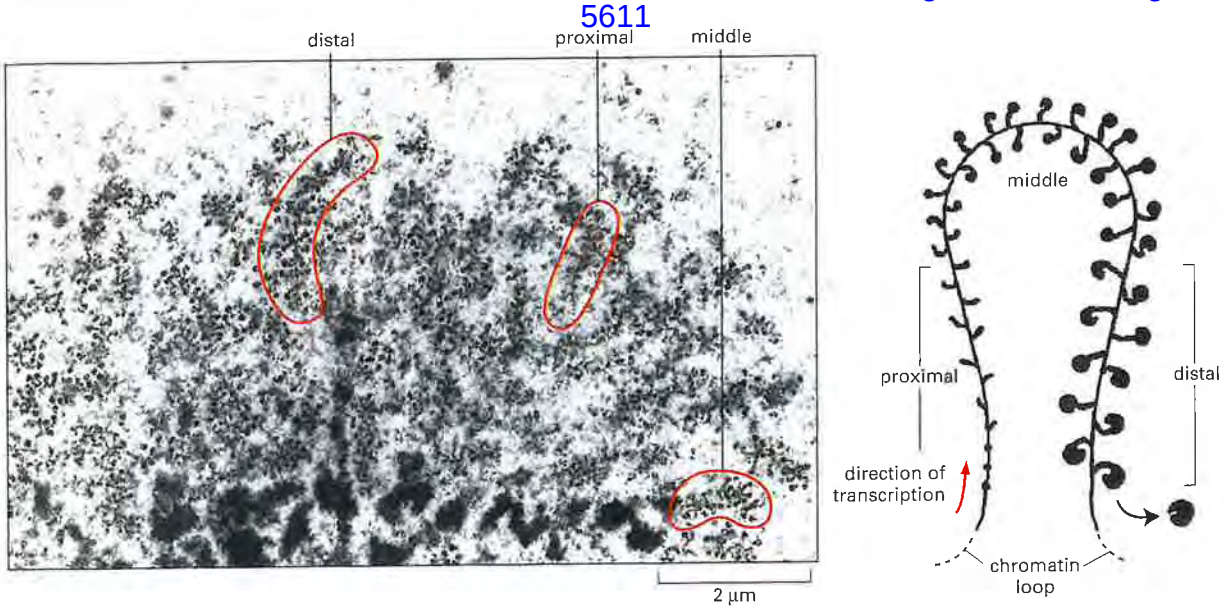


Figure 4–43 Polytene chromosomes from *C. tentans*. The electron micrograph shows a thin section of the chromatin in a Balbiani ring, a chromosome puff very active in gene expression. The Balbiani ring gene codes for secretory proteins the larvae uses to spin a protective tube. The chromatin is arranged in loops, but because the sample has been sectioned, only portions of the loops are visible. As they are synthesized on the chromatin, the RNA molecules are bound up by protein molecules, making them visible as knobs on stalks in the electron microscope. From the size of the RNA–protein complex, the extent of RNA synthesis (transcription) can be inferred; a whole chromatin loop (shown on the right) can then be reconstructed from a set of electron micrograph sections such as that shown here. (Courtesy of B. Daneholt, from U. Skoglund et al., *Cell* 34:847–855, 1983. © Elsevier.)

single gene. When not expressed, the loop of DNA assumes a thickened structure, possibly a folded 30-nm fiber, but when gene expression is occurring, the loop becomes more extended. Both types of loops contain the four core histones and histone H1.

It seems likely that the default loop structure is a folded 30-nm fiber and that the histone modifying enzymes, chromatin remodeling complexes, and other proteins required for gene expression all help to convert it to a more extended form whenever a gene is expressed. In electron micrographs, the chromatin located on either side of the decondensed loop appears considerably more compact, which is consistent with the idea that a loop constitutes an independent functional domain of chromatin structure.

Although controversial, it has been proposed that all of the DNA in polytene chromosomes is arranged in loops that condense and decondense according to when the genes within them are expressed. It may be that all interphase chromosomes from all eucaryotes are also packaged into an orderly series of looped domains, each containing a small number of genes whose expression is regulated in a coordinated way (Figure 4–44). We shall return to this issue in Chapter 7 when we discuss the ways in which gene expression is regulated by the cell.

Heterochromatin Is Highly Organized and Usually Resistant to Gene Expression

Having described some features of interphase chromosomes inferred from a few rare cases, we now turn to characteristics of interphase chromosomes that can be observed in a wide variety of organisms. Light-microscope studies in the 1930s distinguished between two types of chromatin in the interphase nuclei of many higher eucaryotic cells: a highly condensed form, called **heterochromatin**, and all the rest, which is less condensed, called **euchromatin**. Euchromatin is composed of the types of chromosomal structures—30-nm fibers and looped domains—that we have discussed so far. Heterochromatin, in contrast, includes additional proteins and probably represents more compact levels of organization that are just beginning to be understood. In a typical mammalian cell, approximately 10% of the genome is packaged into heterochromatin. Although present in many locations along chromosomes, it is concentrated in specific regions, including the centromeres and telomeres.

Most DNA that is folded into heterochromatin does not contain genes. However, genes that do become packaged into heterochromatin are usually resistant to being expressed, because heterochromatin is unusually compact. This does not mean that heterochromatin is useless or deleterious to the cell; as we see

below, regions of heterochromatin are responsible for the proper functioning of telomeres and centromeres (which lack genes), and its formation may even help protect the genome from being overtaken by "parasitic" mobile elements of DNA. Moreover, a few genes require location in heterochromatin regions if they are to be expressed. In fact, the term *heterochromatin* (which was first defined cytologically) is likely to encompass several distinct types of chromatin structures whose common feature is an especially high degree of organization. Thus, heterochromatin should not be thought of as encapsulating "dead" DNA, but rather as creating different types of compact chromatin with distinct features and roles.

Heterochromatin's resistance to gene expression makes it amenable to study even in organisms in which it cannot be directly observed. When a gene that is normally expressed in euchromatin is experimentally relocated into a region of heterochromatin, it ceases to be expressed, and the gene is said to be *silenced*. These differences in gene expression are examples of **position effects**, in which the activity of a gene depends on its position along a chromosome. First recognized in *Drosophila*, position effects have now been observed in many organisms and they are thought to reflect an influence of the different states of chromatin structure along chromosomes on gene expression. Thus, chromosomes can be considered as mosaics of distinct forms of chromatin, each of which has a special effect on the ability of the DNA it contains to be addressed by the cell.

Many position effects exhibit an additional feature called *position effect variegation*, which is responsible for the mottled appearance of the fly eye and the sectoring of the yeast colony in the examples shown in Figure 4–45. These patterns can result from patches of cells in which a silenced gene has become reactivated; once reactivated, the gene is inherited stably in this form in daughter cells. Alternatively, a gene can start out in euchromatin early in development, and then be selected more or less randomly for packaging into heterochromatin, causing its inactivation in a cell and all of its daughters.

The study of position effect variegation has revealed two important characteristics of heterochromatin. First, heterochromatin is dynamic; it can "spread" into a region and later "retract" from it at low but observable frequencies. Second, the state of chromatin—whether heterochromatin or euchromatin—tends to be inherited from a cell to its progeny. These two features are responsible for position effect variegation, as explained in Figure 4–46. In the next section, we discuss several models to account for the self-sustaining nature of heterochromatin, once it has been formed.

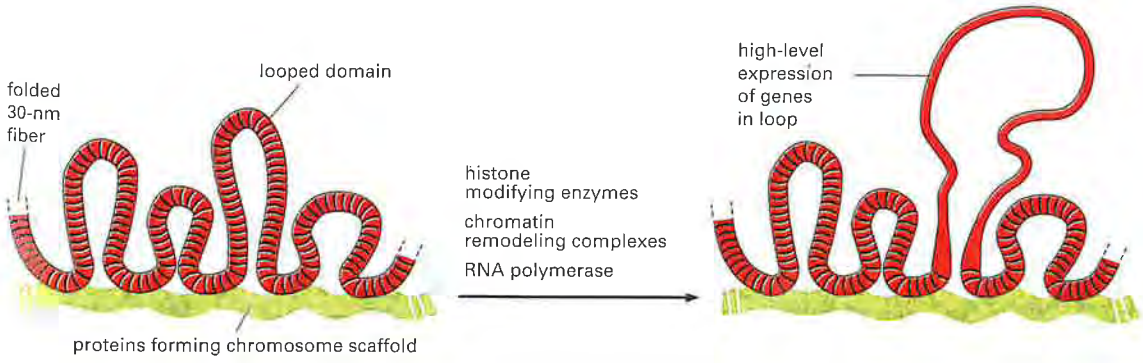


Figure 4–44 A model for the structure of an interphase chromosome. A section of an interphase chromosome is shown folded into a series of looped domains, each containing 20,000–100,000 nucleotide pairs of double-helical DNA condensed into a 30-nm fiber. Individual loops can decondense, perhaps in part through an accordionlike expansion of the 30-nm fiber (see Figure 4–29), when the cell requires direct access to the DNA packaged in these loops. This decondensation is brought about by enzymes that directly modify chromatin structure—as well as by proteins, such as RNA polymerase (discussed in Chapter 6), that act directly on the underlying DNA. It is not understood how the folded 30-nm fiber is anchored to the chromosome axis, but evidence suggests that the base of chromosomal loops is rich in DNA topoisomerases, which are enzymes that allow DNA to swivel when anchored (see pp. 251–253).

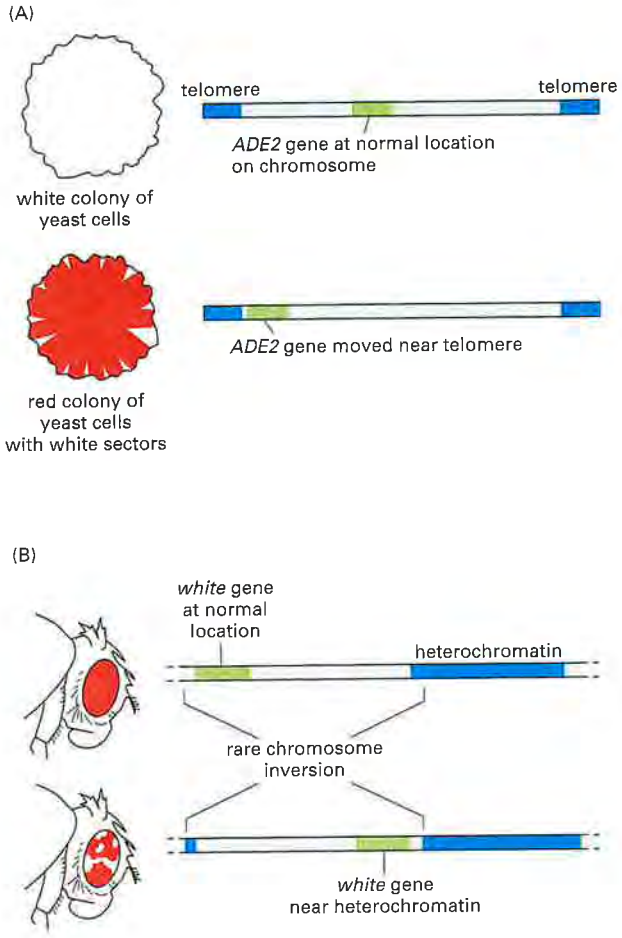


Figure 4–45 Position effects on gene expression in two different eucaryotic organisms. (A) The yeast *ADE2* gene at its normal chromosomal location is expressed in all cells. When moved near the end of a yeast chromosome, which is inferred to be folded into a form of heterochromatin, the gene is no longer expressed in most cells of the population. The *ADE2* gene codes for one of the enzymes of adenine biosynthesis, and the absence of the *ADE2* gene product leads to the accumulation of a red pigment. Therefore, a colony of cells that expresses *ADE2* is *white*, and one composed of cells where the *ADE2* gene is not expressed is *red*. The *white* sectors that fan out from the middle of the *red* colony grown on an agar surface represent the descendants of cells in which the *ADE2* gene has spontaneously become active. These white sectors are thought to result from a heritable change to a less tightly packed state of chromatin near the *ADE2* gene in these cells. Although yeast chromosomes are too small to be seen under the light microscope, the chromatin structure at the ends of yeast chromosomes is thought to have many of the same structural features as the heterochromatin in the chromosomes of larger organisms.

(B) Position effects can also be observed for the *white* gene in the fruit fly *Drosophila*. The *white* gene controls eye pigment production and is named after the mutation that first identified it. Wild-type flies with a normal *white* gene (*white*$^-$) have normal pigment production, which gives them red eyes, but if the *white* gene is mutated and inactivated, the mutant flies (*white*$^-$) make no pigment and have *white* eyes. In flies in which a normal *white*$^+$ gene has been moved near a region of heterochromatin, the eyes are mottled, with both *red* and *white* patches. The *white* patches represent cells in which the *white*$^+$ gene has been silenced by the effects of the heterochromatin. In contrast, the red patches represent cells that express the *white*$^+$ gene because the heterochromatin did not spread across this gene at the time, early in development, when the heterochromatin first formed. As in the yeast, the presence of large patches of *red* and *white* cells indicates that the state of transcriptional activity of the gene is inherited, once determined by its chromatin packaging in the early embryo.

The Ends of Chromosomes Have a Special Form of Heterochromatin

Unlike the nucleosome and the 30-nm fiber, heterochromatin is not well understood structurally. It almost certainly involves an additional level of folding of 30-nm fiber and requires many proteins in addition to the histones. Although its chromosomes are too small to be seen under the light microscope, the molecular nature of heterochromatin is probably best understood in the simple yeast *S. cerevisiae*. Many experiments with yeast cells have shown that the chromatin extending inward roughly 5000 nucleotide pairs from each chromosome end is resistant to gene expression, and probably has a structure that corresponds to at least one type of heterochromatin in the chromosomes of more complex organisms. Extensive genetic analysis has led to the identification of many of the yeast proteins required for this type of gene silencing.

Mutations in any one of a set of yeast *S*ilent *i*nformation *r*egulator (Sir) proteins prevent the silencing of genes located near telomeres, thereby allowing these genes to be expressed. Analysis of these proteins has led to the discovery of a telomere-bound Sir protein complex that recognizes underacetylated N-terminal tails of selected histones (Figure 4–47A). One of the proteins in this complex is a highly conserved histone deacetylase known as Sir2, which has homologs in diverse organisms, including humans, and presumably has a major role in creating a pattern of histone underacetylation unique to heterochromatin. As discussed earlier in this chapter, deacetylation of the histone tails is thought to allow nucleosomes to pack together into tighter arrays and may also render them less susceptible to some chromatin remodeling complexes. In addition, heterochromatin-specific patterns of histone tail modification are likely to attract additional proteins involved in forming and maintaining heterochromatin (see Figure 4–35).

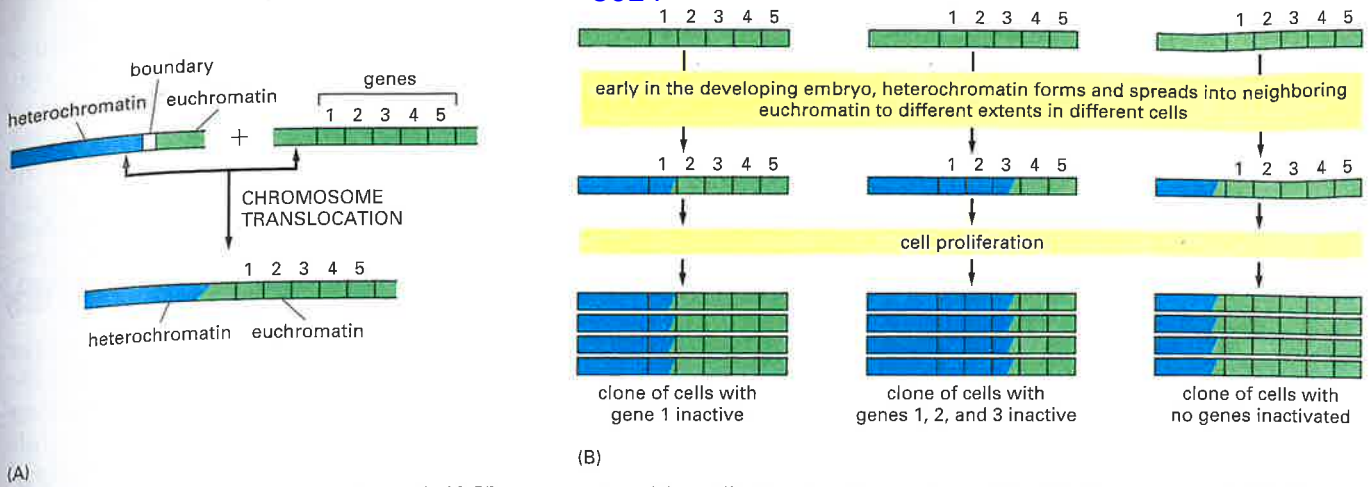


Figure 4–46 The cause of position effect variegation in *Drosophila*. (A) Heterochromatin *(blue)* is normally prevented from spreading into adjacent regions of euchromatin *(green)* by special boundary DNA sequences, which we discuss in Chapter 7. In flies that inherit certain chromosomal rearrangements, however, this barrier is no longer present. (B) During the early development of such flies, heterochromatin can spread into neighboring chromosomal DNA, proceeding for different distances in different cells. This spreading soon stops, but the established pattern of heterochromatin is inherited, so that large clones of progeny cells are produced that have the same neighboring genes condensed into heterochromatin and thereby inactivated (hence the "variegated" appearance of some of these flies; see Figure 4–45B). Although "spreading" is used to describe the formation of new heterochromatin near previously existing heterochromatin, the term may not be wholly accurate. There is evidence that during expansion, heterochromatin can "skip over" some regions of chromatin, sparing the genes that lie within them from repressive effects. One possibility is that heterochromatin can expand across the base of some DNA loops, thus bypassing the chromatin contained in the loop.

But how is the Sir2 protein delivered to the ends of chromosomes in the first place? Another series of experiments has suggested the model shown in Figure 4–47B. A DNA-binding protein that recognizes specific DNA sequences in yeast telomeres also binds to one of the Sir proteins, causing the entire Sir protein complex to assemble on the telomeric DNA. The Sir complex then spreads along the chromosome from this site, modifying the N-terminal tails of adjacent histones

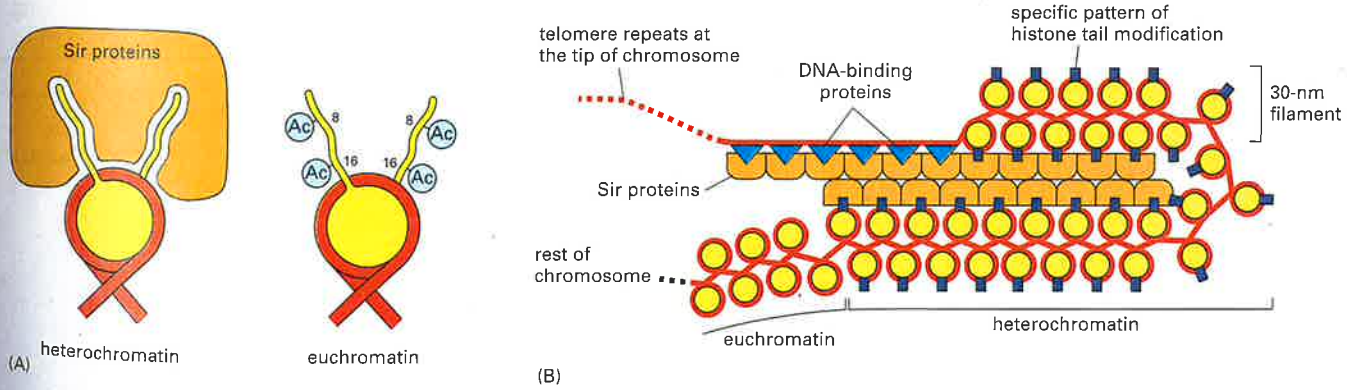


Figure 4–47 Speculative model for the heterochromatin at the ends of yeast chromosomes. (A) Heterochromatin is generally underacetylated, and underacetylated tails of histone H4 are proposed to interact with a complex of Sir proteins, thus stabilizing the association of these proteins with nucleosomes. Although shown as fully unacetylated, the exact pattern of histone H4 tail modification required to bind to the Sir complex is not known with certainty. In some organisms, the methylation of lysine 9 of histone H3 is also a critical signal for heterochromatin formation. In euchromatin, histone tails are typically highly acetylated. Those of H4 are shown as partially acetylated but, in reality, the acetylation state varies across euchromatin. (B) Specialized DNA-binding proteins *(blue triangles)* recognize DNA sequences near the ends of chromosomes and attract the Sir proteins, one of which (Sir2) is an NAD^+-dependent histone deacetylase. This then leads to the cooperative spreading of the Sir protein complex down the chromosome. As this complex spreads, the deacetylation catalyzed by Sir2 helps create new binding sites on nucleosomes for more Sir protein complexes. A "fold back" structure of the type shown may also form.

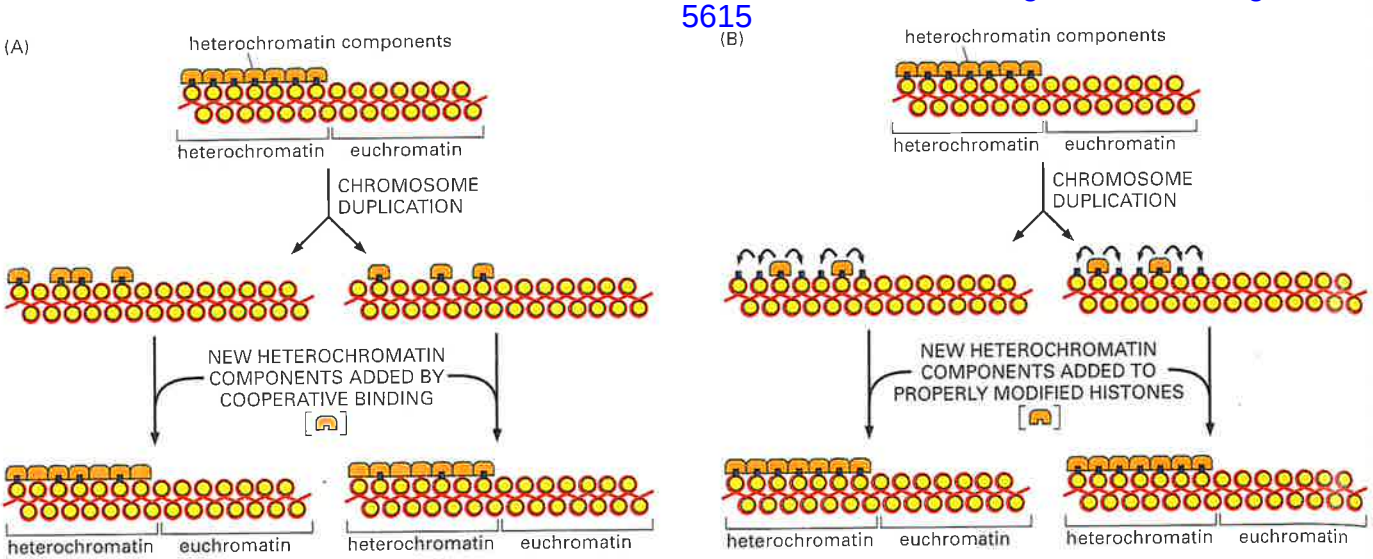


Figure 4–48 Two speculative models for how the tight packaging of DNA in heterochromatin can be inherited during chromosome replication. In both cases, half of the specialized heterochromatin components have been distributed to each daughter chromosome after DNA duplication. (A) In this model, new heterochromatin components bind cooperatively to the inherited components, thereby beginning the process of new heterochromatin formation. The process is completed with the assembly of additional proteins and the eventual covalent modification of the histones (not shown). (B) In this model, the inherited heterochromatin components change the pattern of histone modification on the newly formed daughter nucleosomes nearby, creating new binding sites for free heterochromatin components, which assemble and complete the structure. Both models can account for the spreading effects of heterochromatin, and indeed, both processes may occur simultaneously in cells.

to create the nucleosome-binding sites that the complex prefers. This "spreading effect" is thought to be driven by the cooperative binding of adjacent Sir protein complexes, as well as by the folding back of the chromosome on itself to promote Sir binding in nearby regions (see Figure 4–47B). In addition, the formation of heterochromatin probably requires the action of chromatin remodeling complexes to readjust the positions of nucleosomes as they are packed together.

Unlike most deacetylases, Sir2 requires NAD^+ as a cofactor (see Figure 2–60). The NAD^+ levels in the cell fluctuate with the nutritional health of the cell, increasing as cells become nutritionally deprived. This feature might cause the telomeric heterochromatin to spread in response to starvation (perhaps to silence the expression of genes that are not absolutely required for survival) and then to retract when conditions improve.

The properties of the yeast heterochromatin just described may resemble features of heterochromatin in more complex organisms. Certainly, the spreading of yeast heterochromatin from telomeres is similar in principle to the movement of heterochromatin that causes position effect variegation in animals (see Figure 4–46). Moreover, these properties can be used to explain the heritability of heterochromatin, as outlined in Figure 4–48. Whatever the precise mechanism of heterochromatin formation, it has become clear that covalent modifications of the nucleosome core histones have a critical role in this process. Of special importance in many organisms are the *histone methyl transferases*, enzymes that methylate specific lysines on histones including lysine 9 of histone H3 (see Figure 4–35). This modification is "read" by heterochromatin components (including HP1 in *Drosophila*) that specifically bind this modified form of histone H3 to induce the assembly of heterochromatin. It is likely that a spectrum of different histone modifications is used by the cell to distinguish heterochromatin from euchromatin (see Figure 4–35).

Having the ends of chromosomes packaged into heterochromatin provides several advantages to the cell: it helps to protect the ends of chromosomes from being recognized as broken chromosomes by the cellular repair machinery, it may help to regulate telomere length, and it may assist in the accurate pairing and segregation of chromosomes during mitosis. In Chapter 5 we see that telomeres have additional structural features that distinguish them from other parts of chromosomes.

Centromeres Are Also Packaged into Heterochromatin

Heterochromatin is also observed around centromeres, the DNA sequences that direct the movement of each chromosome into daughter cells every time a cell divides (see Figure 4–22). In many complex organisms, including humans, each

centromere seems to be embedded in a very large stretch of heterochromatin that persists throughout interphase, even though the centromere-directed movement of DNA occurs only during mitosis. The structure and biochemical properties of this so-called *centric heterochromatin* are not well understood, but, like other forms of heterochromatin, it silences the expression of genes that are experimentally placed into it. It contains, in addition to histones (which are typically underacetylated and methylated in heterochromatin), several additional structural proteins that compact the nucleosomes into particularly dense arrangements.

As with telomeres, our best understanding of the chromatin structure of a centromere comes from studies of the much simpler centromeres of the yeast *S. cerevisiae*. Earlier in this chapter we saw that a simple DNA sequence of approximately 125 nucleotide pairs was sufficient to serve as a centromere in this organism. Despite its small size, more than a dozen different proteins assemble on this DNA sequence; the proteins include a histone H3 variant that, along with the other core histones, is believed to form a centromere-specific nucleosome (Figure 4–49A). We do not yet understand what properties this variant type of nucleosome provides to the cell, but similar specialized nucleosomes seem to be present in all eucaryotic centromeres (Figure 4–49B). The additional proteins at the yeast centromere attach it to the spindle microtubules and provide signals that ensure that this attachment is complete before the later stages of mitosis are allowed to proceed (discussed in Chapters 17 and 18).

The centromeres in more complex organisms are considerably larger than those in budding yeasts. For example, fly and human centromeres extend over hundreds of thousands of nucleotide pairs and do not seem to contain a centromere-specific DNA sequence. Rather, most consist largely of short, repeated DNA sequences, known as *alpha satellite DNA* in humans (Figure 4–50). But the same repeat sequences are also found at other (noncentromeric) positions on chromosomes, and how they specify a centromere is poorly understood. Somehow the formation of the inner plate of a kinetochore is "seeded," followed by the cooperative assembly of the entire group of special proteins that form the kinetochore (Figure 4–50B). It seems that centromeres in complex organisms are defined more by an assembly of proteins than by a specific DNA sequence.

There are some striking similarities between the formation and maintenance of centromeres and the formation and maintenance of other regions of heterochromatin. The entire centromere forms as an all-or-none entity, suggesting a highly cooperative addition of proteins after a seeding event. Moreover, once formed, the structure seems to be directly inherited on the DNA as part of each round of chromosome replication. Thus, for example, some regions of our chromosomes contain nonfunctional alpha satellite DNA sequences that seem to be identical to those at the centromere; these sequences are presumed to have arisen from a chromosome-joining event that initially created one chromosome with two centromeres (an unstable, dicentric

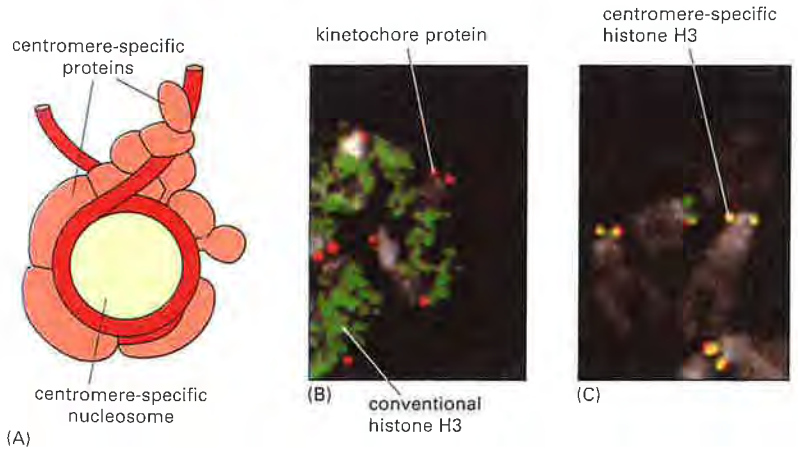


Figure 4–49 The specialized nucleosome formed on centromeres. (A) A model for the proteins that assemble on a yeast centromere. The specialized nucleosome contains an H3 variant (called CENP-A in most organisms), along with core histones H2A, H2B, and H4. The folding of DNA into this nucleosome facilitates the assembly of the other centromere-binding proteins, which form the kinetochore that attaches the centromere to the mitotic spindle (B) The localization of conventional histone H3 on *Drosophila* mitotic chromosomes. The conventional H3 has been fused to a fluorescent protein and appears *green*. A component of the kinetochore has been stained *red* with antibodies against a specific kinetochore protein. (C) The same experiment, but with the centromere-specific histone H3 (instead of the conventional H3) labeled *green*. When the *red* and *green* stains are coincident, the staining appears *yellow*. (A, adapted from P.B. Meluh et al., *Cell* 94:607–613, 1998; B and C, from S. Henikoff et al., *Proc. Natl. Acad. Sci. USA* 97:716–721, 2000. © National Academy of Sciences.)

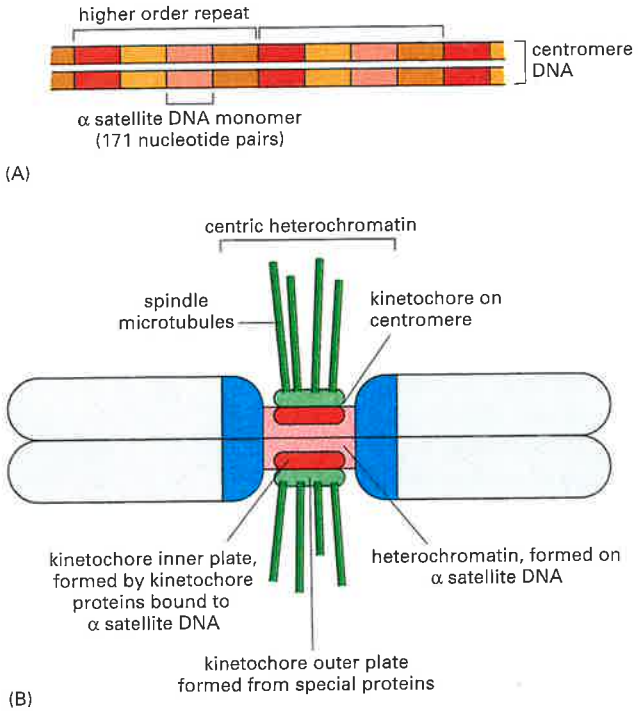


Figure 4–50 The structure of a human centromere. (A) The organization of the alpha satellite DNA sequences, which are repeated many thousands of times at a centromere. (B) An entire chromosome. The alpha satellite DNA sequences *(red)* are AT-rich and consist of a large number of repeats that vary slightly from one another in their DNA sequence. *Blue* represents the position of flanking centric heterochromatin, which contains DNA sequences composed of different types of repeats. As indicated, the kinetochore consists of an inner and an outer plate, formed by a set of kinetochore proteins. The spindle microtubules attach to the kinetochore in M phase of the cell cycle (see Figure 4–22). (B, adapted from T.D. Murphy and G.H. Karpen, *Cell* 93:317–320, 1998.)

chromosome; Figure 4–51A). Moreover, in some unusual cases, new human centromeres (called neocentromeres) have been observed to form spontaneously on fragmented chromosomes. Some of these new positions were originally euchromatic and lack alpha satellite DNA altogether (Figure 4–51B).

To explain these observations it has been proposed that *de novo* centromere formation requires an initial marking (or seeding) event, perhaps the formation of a specialized DNA–protein structure, which, in humans, happens more readily on arrays of alpha satellite DNA than on other DNA sequences. This mark would be duplicated when the chromosome divides, and the same centromere would then function in the next cell division. Very rarely, the mark would be lost after chromosome replication, in which case it would be very difficult to establish again (Figure 4–51C). Although the self-renewing nature of centromeres is not understood in detail, the type of models described for heterochromatin inheritance in Figure 4–48 could also be critical here.

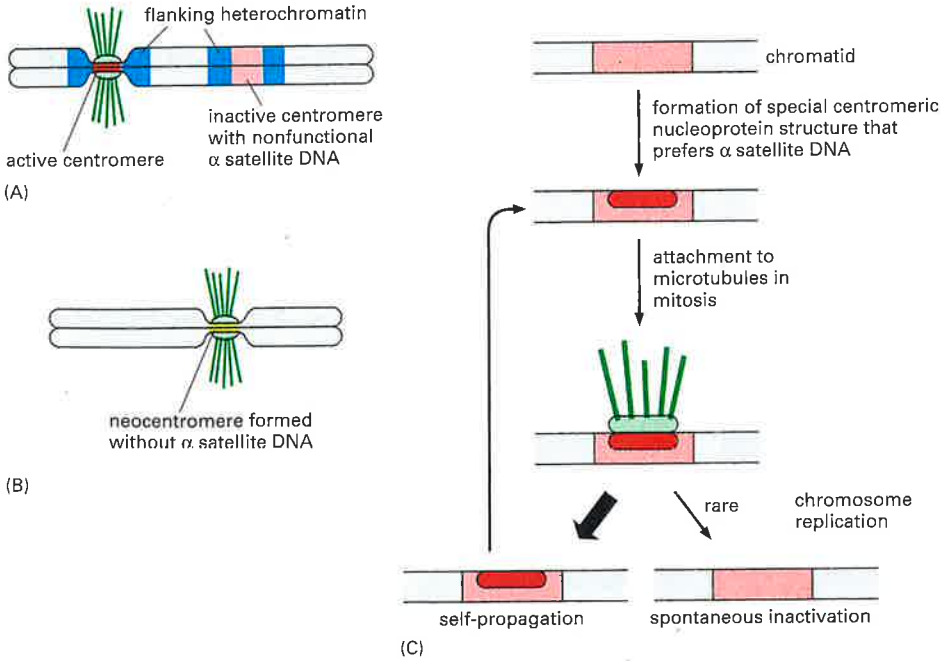


Figure 4–51 The plasticity of human centromere formation. (A) Owing to an ancient chromosome breakage and rejoining event, some human chromosomes contain two blocks of alpha satellite DNA *(red)*, each of which presumably functioned as a centromere in its original chromosome. Usually, these dicentric chromosomes are not stably propagated because they are attached improperly to the spindle and are broken apart during mitosis. In those chromosomes that survive, one of the centromeres has spontaneously inactivated, even though it contains all the necessary DNA sequences. This allows the chromosome to be stably propagated. (B) In a small fraction (1/2000) of human births, extra chromosomes are observed in cells of the offspring. Some of these extra chromosomes, which have formed from a breakage event, lack alpha satellite DNA altogether, yet new centromeres have arisen from what was originally euchromatic DNA. (C) A model to explain the plasticity and inheritance of centromeres.

The plasticity of centromeres may provide an important evolutionary advantage. We have seen that chromosomes evolve in part by breakage and rejoining events (see Figure 4–19). Many of these events produce chromosomes with two centromeres, or chromosome fragments with no centromeres at all. Although rare, the inactivation of centromeres and their ability to be activated *de novo* may occasionally allow newly formed chromosomes to be maintained stably and thereby facilitate the process of chromosome evolution.

Heterochromatin May Provide a Defense Mechanism Against Mobile DNA Elements

DNA packaged in heterochromatin often consists of large tandem arrays of short, repeated sequences that do not code for protein, as we saw above for the heterochromatin of mammalian centromeres. In contrast, euchromatic DNA is rich in genes and other single-copy DNA sequences. Although this correlation is not absolute (some arrays of repeated sequences exist in euchromatin and some genes are present in heterochromatin), this trend suggests that some types of repeated DNA may be a signal for heterochromatin formation. This idea is supported by experiments in which several hundred tandem copies of genes have been artificially introduced into the germ lines of flies and mice. In both organisms these gene arrays are often silenced, and in some cases, they can be observed under the microscope to have formed regions of heterochromatin. In contrast, when single copies of the same genes are introduced into the same position in the chromosome, they are actively expressed.

This feature, called *repeat-induced gene silencing*, may be a mechanism that cells have for protecting their genomes from being overtaken by mobile genetic elements. These elements, which are discussed in Chapter 5, can multiply and insert themselves throughout the genome. According to this idea, once a cluster of such mobile elements has formed, the DNA that contains them would be packaged into heterochromatin to prevent their further proliferation. The same mechanism could be responsible for forming the large regions of heterochromatin that contain large numbers of tandem repeats of a simple sequence, as occurs around centromeres.

Mitotic Chromosomes Are Formed from Chromatin in Its Most Condensed State

Having discussed the dynamic structure of interphase chromosomes, we now turn to the final level of DNA packaging, that observed for mitotic chromosomes. With the exception of a few specialized cases, such as the lampbrush and polytene chromosomes discussed above, most interphase chromosomes are too extended and entangled for their structures to be clearly seen. In contrast, the chromosomes from nearly all eucaryotic cells are readily visible during mitosis, when they coil up to form highly condensed structures. It is remarkable that this further condensation, which reduces the length of a typical interphase chromosome only about tenfold, produces such a dramatic change in the appearance of chromosomes.

Figure 4–52 depicts a typical **mitotic chromosome** at the metaphase stage of mitosis. The two daughter DNA molecules produced by DNA replication during interphase of the cell-division cycle are separately folded to produce two sister chromosomes, or *sister chromatids*, held together at their centromeres (see also Figure 4–21). These chromosomes are normally covered with a variety of molecules, including large amounts of RNA–protein complexes. Once this covering has been stripped away, each chromatid can be seen in electron micrographs to be organized into loops of chromatin emanating from a central scaffolding (Figures 4–53 and 4–54). Several types of experiment demonstrate that the order of visible features along a mitotic chromosome at least roughly reflects the order of the genes along the DNA molecule. Mitotic chromosome condensation can thus be thought of as the final level in the hierarchy of chromosome packaging (Figure 4–55).

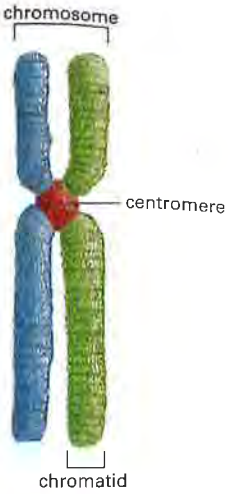


Figure 4–52 A typical mitotic chromosome at metaphase. Each sister chromatid contains one of two identical daughter DNA molecules generated earlier in the cell cycle by DNA replication.

Figure 4–53 A scanning electron micrograph of a region near one end of a typical mitotic chromosome. Each knoblike projection is believed to represent the tip of a separate looped domain. Note that the two identical paired chromatids drawn in Figure 4–52 can be clearly distinguished. (From M.P. Marsden and U.K. Laemmli, *Cell* 17:849–858, 1979. © Elsevier.)

The compaction of chromosomes during mitosis is a highly organized and dynamic process that serves at least two important purposes. First, when condensation is complete (in metaphase), sister chromatids have been disentangled from each other and lie side by side. Thus, the sister chromatids can easily separate when the mitotic apparatus begins pulling them apart. Second, the compaction of chromosomes protects the relatively fragile DNA molecules from being broken as they are pulled to separate daughter cells.

The condensation of interphase chromosomes into mitotic chromosomes occurs in M phase, and it is intimately connected with the progression of the cell cycle, as discussed detail in Chapters 17 and 18. It requires a class of proteins called *condensins* which using the energy of ATP hydrolysis, drive the coiling of each interphase chromosome that produces a mitotic chromosome. Condensins are large protein complexes that contain SMC proteins: long, dimeric protein molecules hinged in the center, with globular domains at each end that bind DNA and hydrolyze ATP (Figure 4–56). When added to purified DNA, condensins use the energy of ATP hydrolysis to make large right-handed loops in the DNA. Although it is not yet known how they act on chromatin, the coiling model shown in Figure 4–56C is based on the fact that condensins are a major structural component of mitotic chromosomes, with one molecule of condensin being present for every 10,000 nucleotides of mitotic DNA.

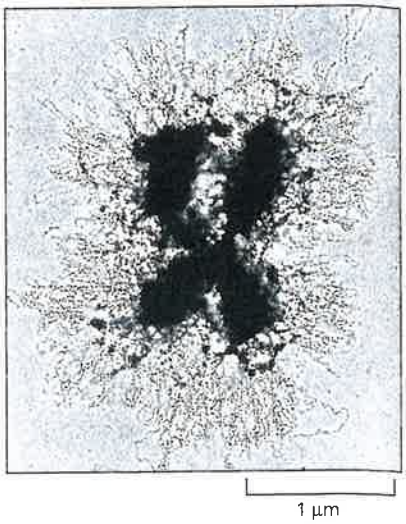


Figure 4–54 An electron micrograph of a mitotic chromosome. This chromosome (from an insect) was treated to reveal loops of chromatin fibers that emanate from a central scaffold of the chromatid. Such micrographs support the idea that the chromatin in all chromosomes is folded into a series of looped domains (see Figure 4–55). (Courtesy of Ulli Laemmli.)

Each Mitotic Chromosome Contains a Characteristic Pattern of Very Large Domains

As mentioned earlier, the display of the 46 human chromosomes at mitosis is called the human karyotype. When stained with dyes such as Giemsa, mitotic

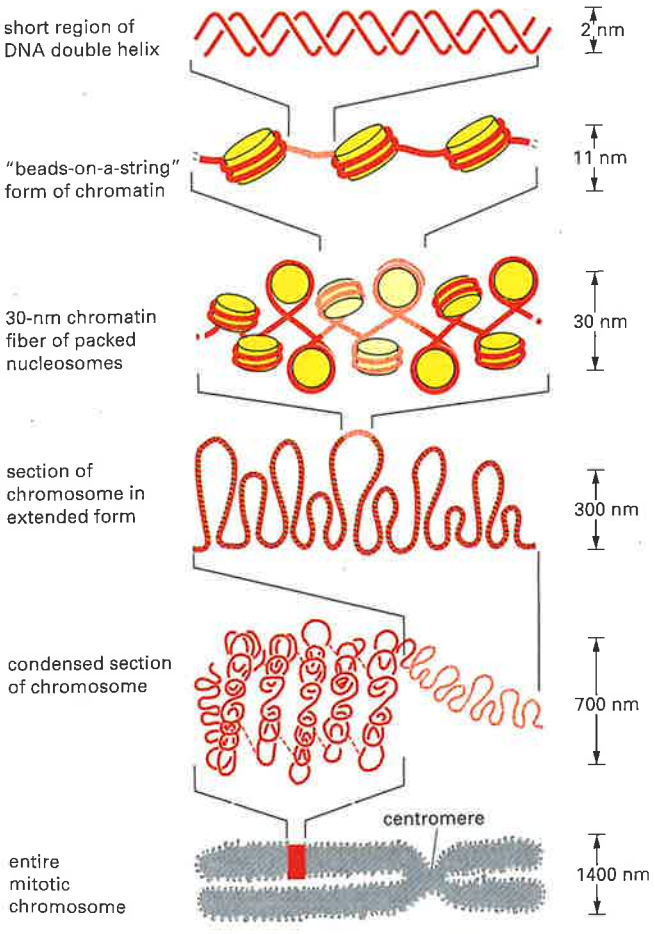


Figure 4–55 Chromatin packing. This model shows some of the many levels of chromatin packing postulated to give rise to the highly condensed mitotic chromosome.

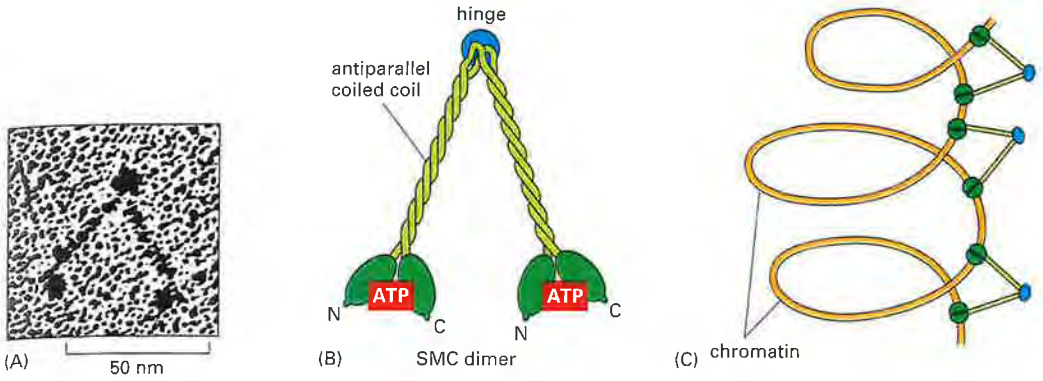


Figure 4–56 The SMC proteins in condensins. (A) Electron micrographs of a purified SMC dimer. (B) The structure of an SMC dimer. The long central region of this protein is an antiparallel coiled coil (see Figure 3–11) with a flexible hinge in its middle, as demonstrated by the electron micrograph in (A). (C) A model for the way in which the SMC proteins in condensins might compact chromatin. In reality, SMC proteins are components of a much larger condensin complex. It has been proposed that, in the cell, condensins coil long strings of looped chromatin domains (see Figure 4–55). In this way the condensins would form a structural framework that maintains the DNA in a highly organized state during M phase of the cell cycle. (A, courtesy of H.P. Erickson; B and C, adapted from T. Hirano, *Genes Dev.* 13:11–19, 1999.)

chromosomes show a striking and reproducible banding pattern along each chromosome, as shown in Figure 4–11. These bands are unrelated to those described earlier for the insect polytene chromosomes, which correspond to relatively small regions of interphase chromatin. In a human mitotic chromosome, all the chromatin is condensed and the bands represent a selective binding of the dyes.

By examining human chromosomes very early in mitosis, when they are less condensed than at metaphase, it has been possible to estimate that the total haploid genome contains about 2000 distinguishable bands. These coalesce progressively as condensation proceeds during mitosis, producing fewer and thicker bands. As we saw earlier, cytogeneticists routinely use the pattern of these chromosome bands to discover in patients genetic alterations such as chromosome inversions, translocations, and other types of chromosomal rearrangements (see Figure 4–12).

Mitotic chromosome bands are detected in chromosomes from species as diverse as humans and flies. Moreover, the pattern of bands in a chromosome has remained unchanged over long periods of evolutionary time. Each human chromosome, for example, has a clearly recognizable counterpart with a nearly identical banding pattern in the chromosomes of the chimpanzee, gorilla, and orangutan—although there are also clear differences, such as chromosome fusion, that give the human 46 chromosomes instead of the ape's 48 (Figure 4–57). This conservation suggests that chromosomes are organized into large domains that may be important for chromosomal function.

Even the thinnest of the bands in Figure 4–11 probably contains more than a million nucleotide pairs, which is nearly the size of a bacterial genome. These bands seem to reflect a rough division of chromosomes into regions of different GC content. The nucleotide sequence of the human genome has revealed large non-random blocks of sequence (some greater than 10^7 nucleotide pairs) that are significantly higher or lower in GC content than the genome-wide average of 41%. The blocks correlate roughly with the staining pattern of metaphase chromosomes. For example, bands that are darkly stained by Giemsa (the so-called G-bands) are correlated with DNA that is low in GC content, whereas lightly stained bands (the R-bands) correspond to DNA of higher than average GC content.

In general, GC-rich regions of the genome have a higher density of genes, especially of "house-keeping" genes, the genes that are expressed in virtually all cell types. On the basis of these observations, it has been proposed that the

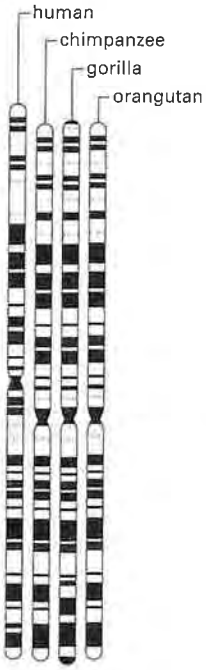


Figure 4–57 Comparison of the Giemsa pattern of the largest human chromosome (chromosome 1) with that of chimpanzee, gorilla, and orangutan. Comparisons among the staining patterns of all the chromosomes indicate that human chromosomes are more closely related to those of chimpanzee than to those of gorilla and that they are more distantly related to those of orangutan. (Adapted from M.W. Strickberger, Evolution, 3rd edn. Sudbury, MA: Jones & Bartlett Publishers, 2000.)

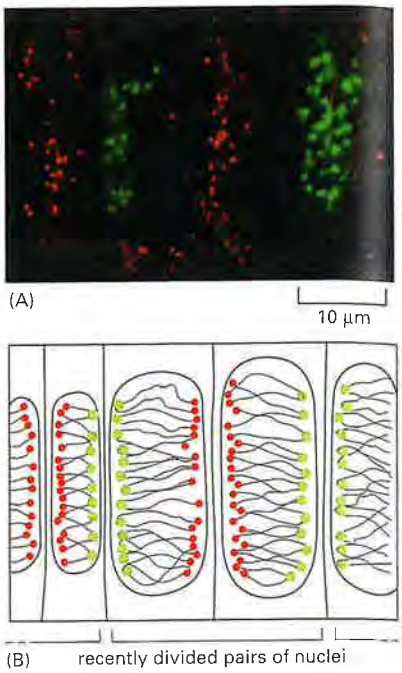


Figure 4–58 The polarized orientation of chromosomes found in certain interphase nuclei. (A) Fluorescent light micrograph of interphase nuclei from the rapidly growing root tip of a plant. Centromeres are stained *green* and teleomeres *red* by *in situ* hybridization of centromere- and telomere-specific DNA sequences coupled to the different fluorescent dyes. (B) Interpretation of (A) showing chromosomes in the Rabl orientation with all the centromeres facing one side of a nucleus and all the telomeres pointing toward the opposite side. (A, from R. Abranches et al., *J. Cell Biol.* 143:5–12, 1998. © The Rockefeller University Press.)

banding pattern may be related to gene expression. Perhaps the differentiation of chromosomes into G- and R-bands reflects subtle differences, determined by GC content, in the way in which chromatin loops are packaged in these areas. If this idea is correct, the rough division of chromosomes can be seen as a form of compartmentalization, in which the particular cellular components involved in gene expression are more concentrated in the R-bands where their activities are required. In any case, it should be obvious from this discussion that we are only beginning to glimpse the principles of large-scale chromosome organization.

Individual Chromosomes Occupy Discrete Territories in an Interphase Nucleus

We saw earlier in this chapter that chromosomes from eucaryotes are contained in the cell nucleus. However, the nucleus is not simply a bag of chromosomes; rather, the chromosomes—as well as the other components inside the nucleus which we shall encounter in subsequent chapters—are highly organized. The way in which chromosomes are organized in the nucleus during interphase, when they are active and difficult to see, has intrigued biologists since the nineteenth century. Although our understanding today is far from complete, we do know some interesting features of these chromosome arrangements.

A certain degree of chromosomal order results from the configuration that the chromosomes always have at the end of mitosis. Just before a cell divides, the condensed chromosomes are pulled to each spindle pole by microtubules attached to the centromeres; thus, as the chromosomes move, the centromeres lead the way and the distal arms (terminating in the telomeres) lag behind. The chromosomes in some nuclei tend to retain this so-called *Rabl orientation* throughout interphase, with their centromeres facing one pole of the nucleus and their telomeres pointing toward the opposite pole (Figures 4–58 and 4–59).

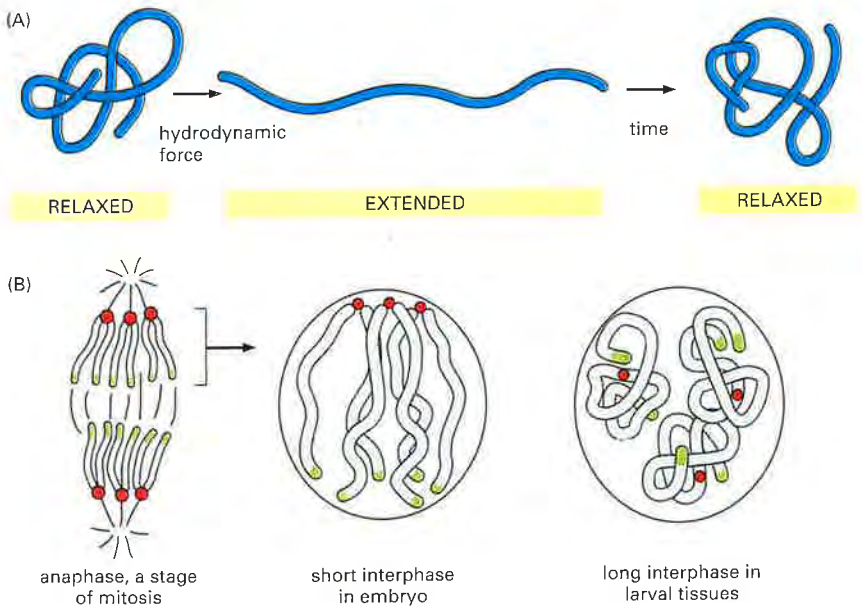


Figure 4–59 A polymer analogy for interphase chromosome organization. (A) The behavior of a polymer in solution. Entropy drives a long polymer into a compact conformation in the absence of an externally applied force. If the polymer is subjected to shear or hydrodynamic force, it becomes extended. But once the force is removed, the polymer chain returns to a more favorable, compact conformation. (B) The behavior of interphase chromosomes may reflect the same simple principles. In *Drosophila* embryos, for example, mitotic divisions occur at intervals of about 10 minutes; during the short intervening interphases, the chromosomes have little time to relax from the Rabl orientation induced by their movement during mitosis. However, in later stages of development, when interphase is much longer, the chromosomes have time to fold up. This folding may be strongly affected by specific associations between different regions of the same chromosome. (Adapted from A.F. Dernburg et al., *Cell* 85:745–759, 1996.)

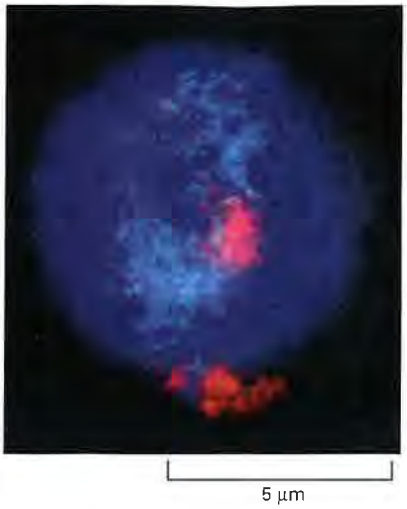


Figure 4–60 Selective "painting" of two interphase chromosomes in a human peripheral lymphocyte. The fluorescent light micrograph shows that the two copies of human chromosome 18 *(red)* and chromosome 19 *(turquoise)* occupy discrete territories of the nucleus. (From J.A. Croft et al., *J. Cell Biol.* 145:1119–1131, 1999. © The Rockefeller University Press.)

The chromosomes in most interphase cells are not found in the Rabl orientation; instead, the centromeres seem to be dispersed in the nucleus. Most cells have a longer interphase than the specialized cells illustrated above, and this presumably gives their chromosomes time to assume a different conformation (see Figure 4–59). Nevertheless, each interphase chromosome does tend to occupy a discrete and relatively small territory in the nucleus: that is, the different chromosomes are not extensively intertwined (Figure 4–60).

One device for organizing chromosomes in the nucleus may be the attachment of certain portions of each chromosome to the nuclear envelope (Figure 4–61). For example, in many cells, telomeres seem bound in this way. But the exact position of a chromosome in a nucleus is not fixed. In the same tissue, for example, two apparently identical cells can have different chromosomes as nearest neighbors.

Some cell biologists believe that there is an intranuclear framework, analogous to the cytoskeleton, on which chromosomes and other components of the nucleus are organized. The *nuclear matrix*, or *scaffold*, has been defined as the insoluble material left in the nucleus after a series of biochemical extraction steps. Some of the proteins that constitute it can be shown to bind specific DNA sequences called *SARs* or *MARs* (scaffold-associated or matrix-associated regions). These DNA sequences have been postulated to form the base of chromosomal loops (see Figure 4–44), or to attach chromosomes to the nuclear envelope and other structures in the nucleus. By means of such chromosomal attachment sites, the matrix might help to organize chromosomes, localize genes, and regulate gene expression and DNA replication. It still remains uncertain, however, whether the matrix isolated by cell biologists represents a structure that is present in intact cells.

Summary

Chromosomes are generally decondensed during interphase, so that their structure is difficult to visualize directly. Notable exceptions are the specialized lampbrush chromosomes of vertebrate oocytes and the polytene chromosomes in the giant secretory cells of insects. Studies of these two types of interphase chromosomes suggest that each long DNA molecule in a chromosome is divided into a large number of discrete domains organized as loops of chromatin, each loop probably consisting of a folded 30-nm chromatin fiber. When genes contained in a loop are expressed, the loop decondenses and allows the cell's machinery easy access to the DNA.

Euchromatin makes up most of interphase chromosomes and probably corresponds to looped domains of 30-nm fibers. However, euchromatin is interrupted by stretches of heterochromatin, in which 30-nm fibers are subjected to additional levels of packing that usually render it resistant to gene expression. Heterochromatin is commonly found around centromeres and near telomeres, but it is also present at other positions on chromosomes. Although considerably less condensed than mitotic chromosomes, interphase chromosomes occupy discrete territories in the cell nucleus; that is, they are not extensively intertwined.

All chromosomes adopt a highly condensed conformation during mitosis. When they are specially stained, these mitotic chromosomes have a banding structure that allows each individual chromosome to be recognized unambiguously. These bands contain millions of DNA nucleotide pairs, and they reflect a poorly-understood coarse heterogeneity of chromosome structure.

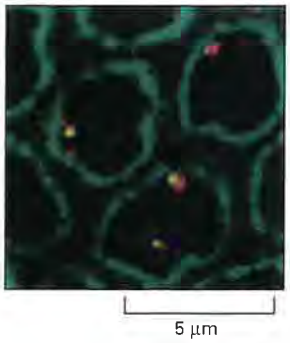


Figure 4–61 Specific regions of interphase chromosomes in close proximity to the nuclear envelope. This high-resolution microscopic view of nuclei from a *Drosophila* embryo shows the localization of two different regions of chromosome 2 *(yellow* and *magenta)* close to the nuclear envelope (stained green with antilamina antibodies). Other regions of the same chromosome are more distant from the envelope. (From W.F. Marshall et al., *Mol. Biol. Cell* 7:825–842, 1996.)

References

General

Hartwell L, Hood L, Goldberg ML et al. (2000) Genetics: from Genes to Genomes. Boston: McGraw Hill.

Lewin B (2000) Genes VII. Oxford: Oxford University Press.

Lodish H, Berk A, Zipursky SL et al. (2000) Molecular Cell Biology, 4th edn. New York: WH Freeman.

Wolfe A (1999) Chromatin: Structure and Function, 3rd edn. New York: Academic Press.

The Structure and Function of DNA

Avery OT, MacLeod CM & McCarty M (1944) Studies on the chemical nature of the substance inducing transformation of pneumococcal types. *J. Exp. Med.* 79, 137.

Meselson M & Stahl FW (1958) The replication of DNA in *E. coli*. *Proc. Natl Acad. Sci. USA* 44, 671–682.

Watson JD & Crick FHC (1953) Molecular structure of nucleic acids. A structure for deoxyribose nucleic acids. *Nature* 171, 737–738.

Chromosomal DNA and Its Packaging in the Chromatin Fiber

Aalfs JD & Kingston RE (2000) What does 'chromatin remodeling' mean? *Trends Biochem. Sci.* 25, 548–555.

Cairns BR (1998) Chromatin remodeling machines: similar motors, ulterior motives. *Trends Biochem. Sci.* 23, 20–25.

Carter NP (1994) Cytogenetic analysis by chromosome painting. *Cytometry* 18, 2–10.

Cheung P, Allis CD & Sassone-Corsi P (2000) Signaling to chromatin through histone modifications. *Cell* 103, 263–271.

Clark MS (1999) Comparative genomics: the key to understanding the Human Genome Project. *Bioessays* 21, 121–130.

DePamphilis ML (1999) Replication origins in metazoan chromosomes: fact or fiction? *Bioessays* 21, 5–16.

Dunham I, Shimizu N, Roe BA et al. (1999) The DNA sequence of human chromosome 22. *Nature* 402, 489–495.

Felsenfeld G (1985) DNA. *Sci. Am.* 253(4), 58–67.

Grunstein M (1992) Histones as regulators of genes. *Sci. Am.* 267(4), 68–74B.

International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature* 409, 860–921.

Jenuwein T & Allis CD (2001) Translating the histone code. *Science* 293:1074–1080.

Kingston RE & Narlikar GJ (1999) ATP-dependent remodeling and acetylation as regulators of chromatin fluidity. *Genes Dev.* 13, 2339–2352.

Kornberg RD & Lorch Y (1999) Chromatin-modifying and -remodeling complexes. *Curr. Opin. Genet. Dev.* 9, 148–151.

Kornberg RD & Lorch Y (1999) Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell* 98, 285–294.

Luger K & Richmond TJ (1998) The histone tails of the nucleosome. *Curr. Opin. Genet. Dev.* 8, 140–146.

Luger K, Mader AW, Richmond RK et al. (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389, 251–260.

McEachern MJ, Krauskopf A & Blackburn EH (2000) Telomeres and their control. *Annu. Rev. Genet.* 34, 331–358.

Ng HH & Bird A (2000) Histone deacetylases: silencers for hire. *Trends Biochem. Sci.* 25, 121–126.

O'Brien S, Menotti-Raymond M, Murphy W et al. (1999) The promise of comparative genomics in mammals. *Science* 286, 458–480.

Pidoux AL & Allshire RC (2000) Centromeres: getting a grip of chromosomes. *Curr. Opin. Cell Biol.* 12, 308–319.

Rhodes D (1997) Chromatin structure. The nucleosome core all wrapped up. *Nature* 389, 231–233.

Rice JC & Allis CD (2001) Histone methylation versus histone acetylation: new insights into epigenetic regulation. *Curr. Opin. Cell Biol.* 13, 263–273.

Ried T, Schrock E, Ning Y & Wienberg J (1998) Chromosome painting: a useful art. *Hum. Mol. Genet.* 7, 1619–1626.

Rubin GM (2001) Comparing species. *Nature* 409, 820–821.

Stewart A (1990) The functional organization of chromosomes and the nucleus—a special issue. *Trends Genet.* 6, 377–379.

Strahl BD & Allis CD (2000) The language of covalent histone modifications. *Nature* 403, 41–45.

Travers AA (1987) DNA bending and nucleosome positioning. *Trends Biochem. Sci.* 12, 108–112.

Wu J & Grunstein M (2000) 25 years after the nucleosome model: chromatin modifications. *Trends Biochem. Sci.* 25, 619–623.

The Global Structure of Chromosomes

Agard DA & Sedat JW (1983) Three-dimensional architecture of a polytene nucleus. *Nature* 302, 676–681.

Ashburner M, Chihara C, Meltzer P & Richards G (1974) Temporal control of puffing activity in polytene chromosomes. *Cold Spring Harbor Symp. Quant. Biol.* 38, 655–662.

Bickmore WA & Sumner AT (1989) Mammalian chromosome banding—an expression genome organization. *Trends Genet.* 5, 144–148.

Birchler JA, Bhadra MP & Bhadra U (2000) Making noise about silence: repression of repeated genes in animals. *Curr. Opin. Genet. Dev.* 10, 211–216.

Callan HG (1982) Lampbrush chromosomes. *Proc. Roy. Soc. Lond. Ser. B. (Biol.)* 214, 417–448.

Croft JA, Bridger JM, Boyle S et al. (1999) Differences in the localization and morphology of chromosomes in the human nucleus. *J. Cell Biol.* 145, 1119–1131.

Griffith JD, Comeau L, Rosenfield S et al. (1999) Mammalian telomeres end in a large duplex loop. *Cell* 97, 503–514.

Grunstein M (1997) Molecular model for telomeric heterochromatin in yeast. *Curr. Opin. Cell Biol.* 9, 383–387.

Hart CM & Laemmli UK (1998) Facilitation of chromatin dynamics by SARs. *Curr. Opin. Genet. Dev.* 8, 519–525.

Henikoff S (1990) Position-effect variegation after 60 years. *Trends Genet.* 6, 422–426.

Henikoff S (1998) Conspiracy of silence among repeated transgenes. *Bioessays* 20, 532–535.

Hirano T (1999) SMC-mediated chromosome mechanics: a conserved scheme from bacteria to vertebrates. *Genes Dev.* 13, 11–19.

Hirano T (2000) Chromosome cohesion, condensation, and separation. *Annu. Rev. Biochem.* 69, 115–144.

Lamond AI & Earnshaw WC (1998) Structure and function in the nucleus. *Science* 280, 547–553.

Lyko F & Paro R (1999) Chromosomal elements conferring epigenetic inheritance. *Bioessays* 21, 824–832.

Marsden M & Laemmli UK (1979) Metaphase chromosome structure: evidence for a radial loop model. *Cell* 17, 849–858.

Pluta AF, Mackay AM, Ainsztein AM et al. (1995) The centromere: hub of chromosomal activities. *Science* 270, 1591–1594.

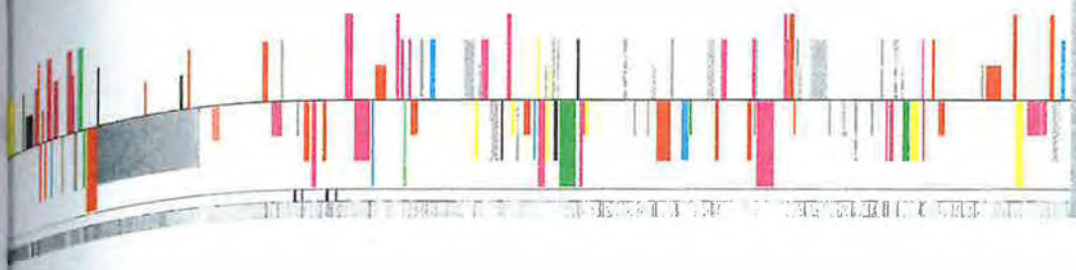
Saitoh N, Goldberg I & Earnshaw WC (1995) The SMC proteins and the coming of age of the chromosome scaffold hypothesis. *Bioessays* 17, 759–766.

Spector DL (1993) Macromolecular domains within the cell nucleus. *Annu. Rev. Cell Biol.* 9, 265–315.

Thummel CS (1990) Puffs and gene regulation—molecular insights into the Drosophila ecdysone regulatory hierarchy. *Bioessays* 12, 561–568.

Weiler KS & Wakimoto BT (1995) Heterochromatin and gene expression in Drosophila. *Annu. Rev. Genet.* 29, 577–605.

Zhimulev IF (1998) Morphology and structure of polytene chromosomes. *Adv. Genet.* 37, 1–566.



6

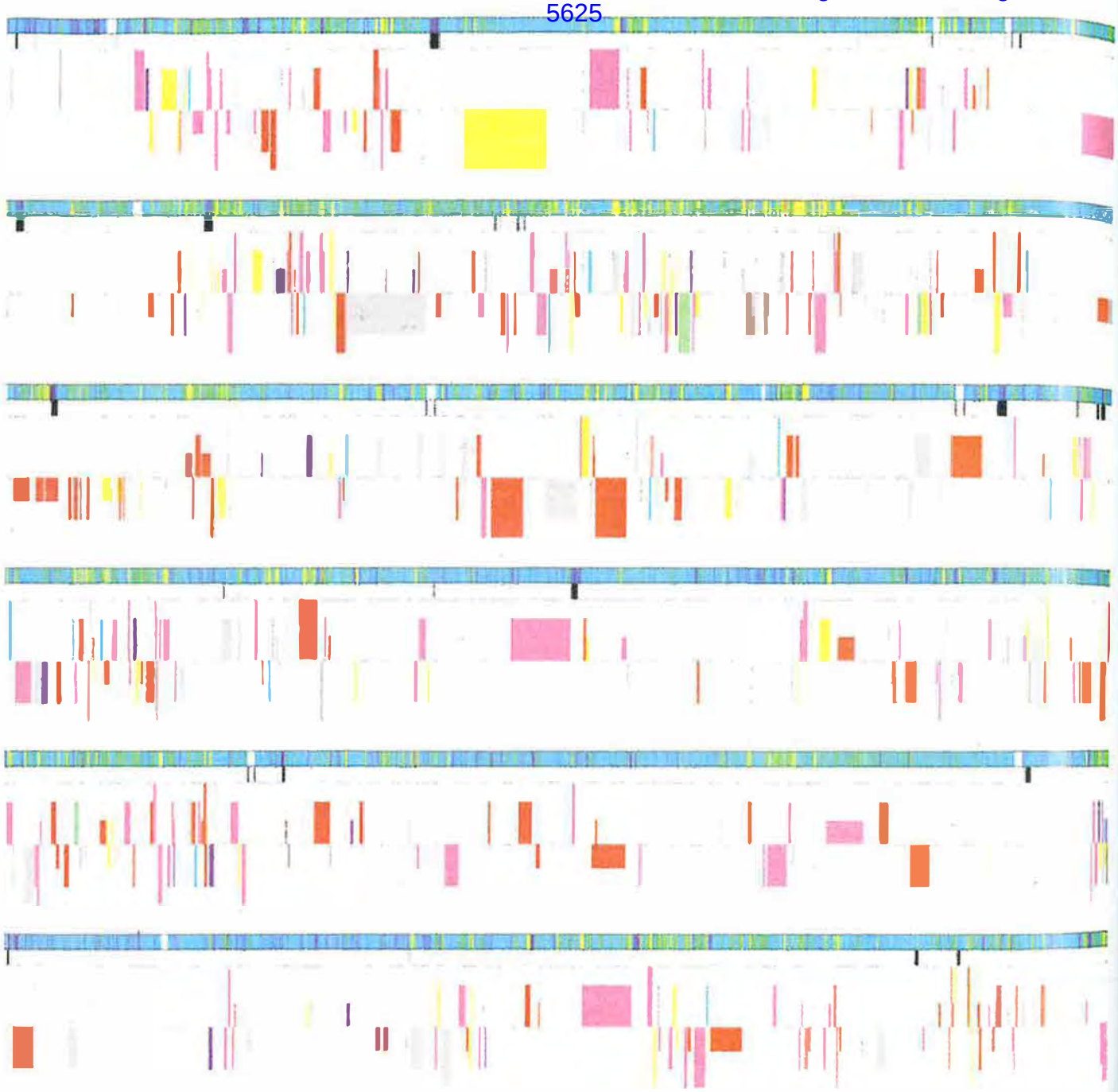
HOW CELLS READ THE GENOME: FROM DNA TO PROTEIN

Only when the structure of DNA was discovered in the early 1950s did it become clear how the hereditary information in cells is encoded in DNA's sequence of nucleotides. The progress since then has been astounding. Fifty years later, we have complete genome sequences for many organisms, including humans, and we therefore know the maximum amount of information that is required to produce a complex organism like ourselves. The limits on the hereditary information needed for life constrain the biochemical and structural features of cells and make it clear that biology is not infinitely complex.

In this chapter, we explain how cells decode and use the information in their genomes. We shall see that much has been learned about how the genetic instructions written in an alphabet of just four "letters"—the four different nucleotides in DNA—direct the formation of a bacterium, a fruitfly, or a human. Nevertheless, we still have a great deal to discover about how the information stored in an organism's genome produces even the simplest unicellular bacterium with 500 genes, let alone how it directs the development of a human with approximately 30,000 genes. An enormous amount of ignorance remains; many fascinating challenges therefore await the next generation of cell biologists.

The problems cells face in decoding genomes can be appreciated by considering a small portion of the genome of the fruit fly *Drosophila melanogaster* (Figure 6–1). Much of the DNA-encoded information present in this and other genomes is used to specify the linear order—the sequence—of amino acids for every protein the organism makes. As described in Chapter 3, the amino acid sequence in turn dictates how each protein folds to give a molecule with a distinctive shape and chemistry. When a particular protein is made by the cell, the corresponding region of the genome must therefore be accurately decoded. Additional information encoded in the DNA of the genome specifies exactly when in the life of an organism and in which cell types each gene is to be expressed into protein. Since proteins are the main constituents of cells, the decoding of the genome determines not only the size, shape, biochemical properties, and behavior of cells, but also the distinctive features of each species on Earth.

One might have predicted that the information present in genomes would be arranged in an orderly fashion, resembling a dictionary or a telephone directory.



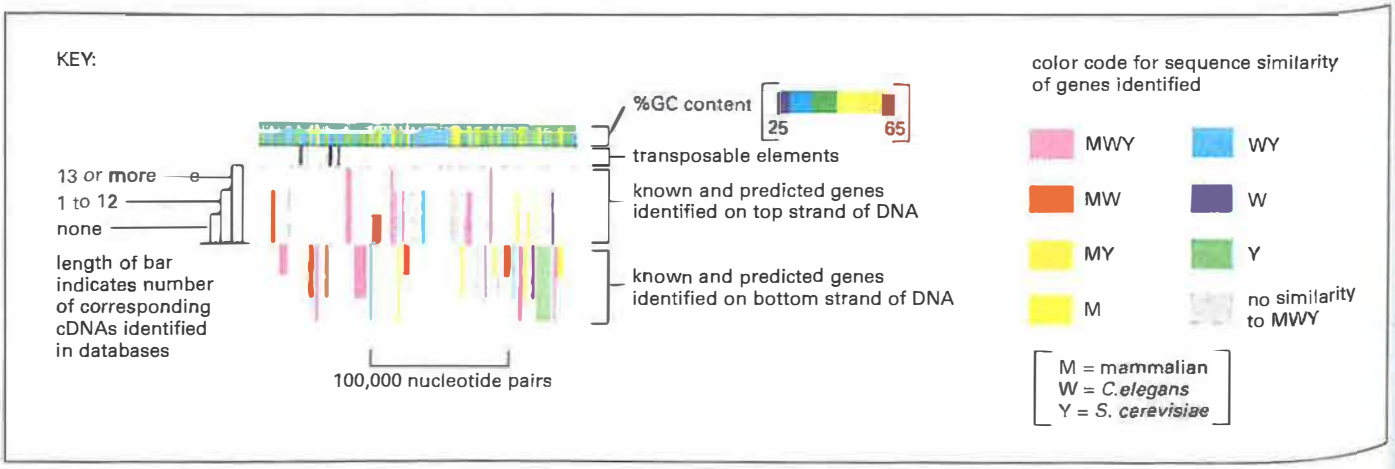

KEY:
%GC content
25
65
transposable elements
known and predicted genes identified on top strand of DNA
known and predicted genes identified on bottom strand of DNA
13 or more
1 to 12
none
length of bar indicates number of corresponding cDNAs identified in databases
100,000 nucleotide pairs
color code for sequence similarity of genes identified
MWY
WY
MW
W
MY
Y
M
no similarity to MWY
M = mammalian
W = C.elegans
Y = S. cerevisiae

Figure 6–1 *(opposite page)* **Schematic depiction of a portion of chromosome 2 from the genome of the fruit fly *Drosophila melanogaster*.** This figure represents approximately 3% of the total *Drosophila* genome, arranged as six contiguous segments. As summarized in the key, the symbolic representations are: *rainbow-colored bar*: G–C base-pair content; *black vertical lines* of various thicknesses: locations of transposable elements, with thicker bars indicating clusters of elements; *colored boxes*: genes (both known and predicted) coded on one strand of DNA (boxes *above* the midline) and genes coded on the other strand (boxes *below* the midline). The length of each predicted gene includes both its exons (protein-coding DNA) and its introns (non-coding DNA) (see Figure 4–25). As indicated in the key, the height of each gene box is proportional to the number of cDNAs in various databases that match the gene. As described in Chapter 8, cDNAs are DNA copies of mRNA molecules, and large collections of the nucleotide sequences of cDNAs have been deposited in a variety of databases. The higher the number of matches between the nucleotide sequences of cDNAs and that of a particular predicted gene, the higher the confidence that the predicted gene is transcribed into RNA and is thus a genuine gene. The color of each gene box (see *color code* in the key) indicates whether a closely related gene is known to occur in other organisms. For example, MWY means the gene has close relatives in mammals, in the nematode worm *Caenorhabditis elegans*, and in the yeast *Saccharomyces cerevisiae*. MW indicates the gene has close relatives in mammals and the worm but not in yeast. (From Mark D. Adams et al., *Science* 287:2185–2195, 2000. © AAAS.)

Although the genomes of some bacteria seem fairly well organized, the genomes of most multicellular organisms, such as our *Drosophila* example, are surprisingly disorderly. Small bits of coding DNA (that is, DNA that codes for protein) are interspersed with large blocks of seemingly meaningless DNA. Some sections of the genome contain many genes and others lack genes altogether. Proteins that work closely with one another in the cell often have their genes located on different chromosomes, and adjacent genes typically encode proteins that have little to do with each other in the cell. Decoding genomes is therefore no simple matter. Even with the aid of powerful computers, it is still difficult for researchers to locate definitively the beginning and end of genes in the DNA sequences of complex genomes, much less to predict when each gene is expressed in the life of the organism. Although the DNA sequence of the human genome is known, it will probably take at least a decade for humans to identify every gene and determine the precise amino acid sequence of the protein it produces. Yet the cells in our body do this thousands of times a second.

The DNA in genomes does not direct protein synthesis itself, but instead uses RNA as an intermediary molecule. When the cell needs a particular protein, the nucleotide sequence of the appropriate portion of the immensely long DNA molecule in a chromosome is first copied into RNA (a process called *transcription*). It is these RNA copies of segments of the DNA that are used directly as templates to direct the synthesis of the protein (a process called *translation*). The flow of genetic information in cells is therefore from DNA to RNA to protein (Figure 6–2). All cells, from bacteria to humans, express their genetic information in this way—a principle so fundamental that it is termed the *central dogma* of molecular biology.

Despite the universality of the central dogma, there are important variations in the way information flows from DNA to protein. Principal among these is that RNA transcripts in eucaryotic cells are subject to a series of processing steps in the nucleus, including *RNA splicing*, before they are permitted to exit from the nucleus and be translated into protein. These processing steps can critically change the "meaning" of an RNA molecule and are therefore crucial for understanding how eucaryotic cells read the genome. Finally, although we focus on the production of the proteins encoded by the genome in this chapter, we see that for some genes RNA is the final product. Like proteins, many of these RNAs fold into precise three-dimensional structures that have structural and catalytic roles in the cell.

We begin this chapter with the first step in decoding a genome: the process of transcription by which an RNA molecule is produced from the DNA of a gene. We then follow the fate of this RNA molecule through the cell, finishing when a correctly folded protein molecule has been formed. At the end of the chapter, we consider how the present, quite complex, scheme of information storage, transcription, and translation might have arisen from simpler systems in the earliest stages of cellular evolution.

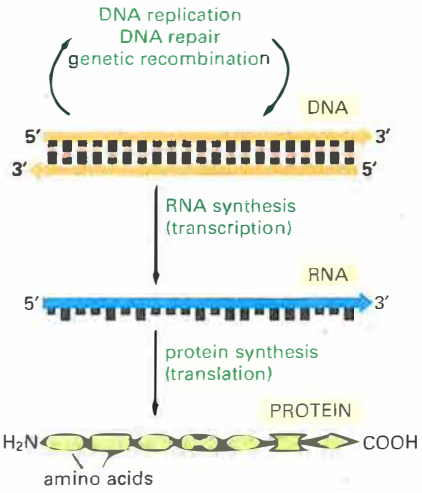


Figure 6–2 The pathway from DNA to protein. The flow of genetic information from DNA to RNA (transcription) and from RNA to protein (translation) occurs in all living cells.

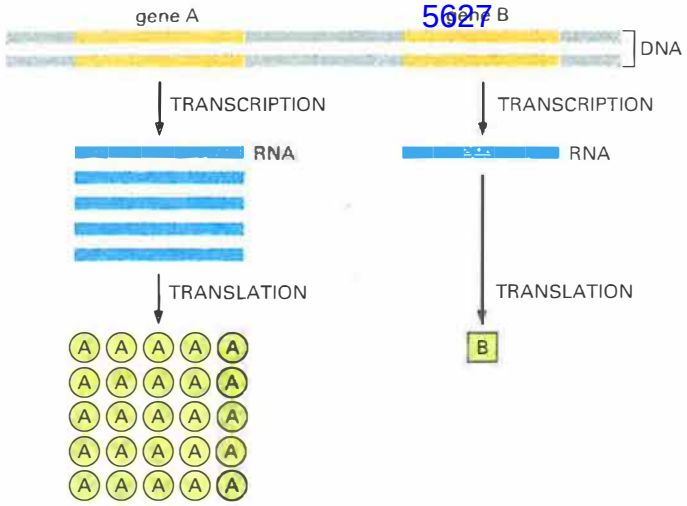


Figure 6–3 Genes can be expressed with different efficiencies. Gene A is transcribed and translated much more efficiently than gene B. This allows the amount of protein A in the cell to be much greater than that of protein B.

FROM DNA TO RNA

Transcription and translation are the means by which cells read out, or express, the genetic instructions in their genes. Because many identical RNA copies can be made from the same gene, and each RNA molecule can direct the synthesis of many identical protein molecules, cells can synthesize a large amount of protein rapidly when necessary. But each gene can also be transcribed and translated with a different efficiency, allowing the cell to make vast quantities of some proteins and tiny quantities of others (Figure 6–3). Moreover, as we see in the next chapter, a cell can change (or regulate) the expression of each of its genes according to the needs of the moment—most obviously by controlling the production of its RNA.

Portions of DNA Sequence Are Transcribed into RNA

The first step a cell takes in reading out a needed part of its genetic instructions is to copy a particular portion of its DNA nucleotide sequence—a gene—into an RNA nucleotide sequence. The information in RNA, although copied into another chemical form, is still written in essentially the same language as it is in DNA—the language of a nucleotide sequence. Hence the name **transcription**.

Like DNA, RNA is a linear polymer made of four different types of nucleotide subunits linked together by phosphodiester bonds (Figure 6–4). It differs from DNA chemically in two respects: (1) the nucleotides in RNA are *ribonucleotides*—that is, they contain the sugar ribose (hence the name *ribo*nucleic acid) rather than deoxyribose; (2) although, like DNA, RNA contains the bases adenine (A), guanine (G), and cytosine (C), it contains the base uracil (U) instead of the thymine (T) in DNA. Since U, like T, can base-pair by hydrogen-bonding with A (Figure 6–5), the complementary base-pairing properties described for DNA in Chapters 4 and 5 apply also to RNA (in RNA, G pairs with C, and A pairs with U). It is not uncommon, however, to find other types of base pairs in RNA: for example, G pairing with U occasionally.

Despite these small chemical differences, DNA and RNA differ quite dramatically in overall structure. Whereas DNA always occurs in cells as a double-stranded helix, RNA is single-stranded. RNA chains therefore fold up into a variety of shapes, just as a polypeptide chain folds up to form the final shape of a protein (Figure 6–6). As we see later in this chapter, the ability to fold into complex three-dimensional shapes allows some RNA molecules to have structural and catalytic functions.

Transcription Produces RNA Complementary to One Strand of DNA

All of the RNA in a cell is made by DNA transcription, a process that has certain similarities to the process of DNA replication discussed in Chapter 5.

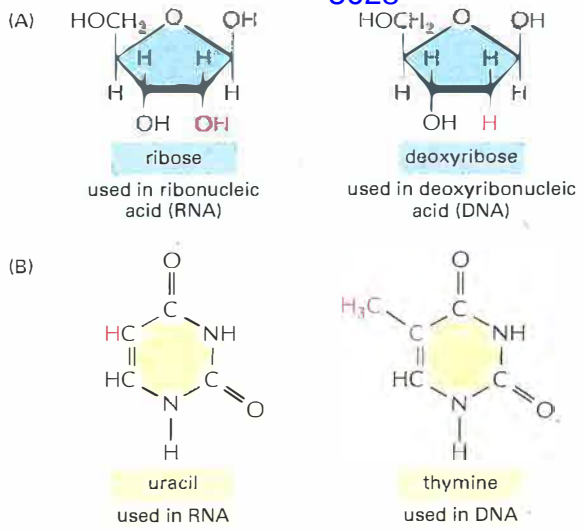
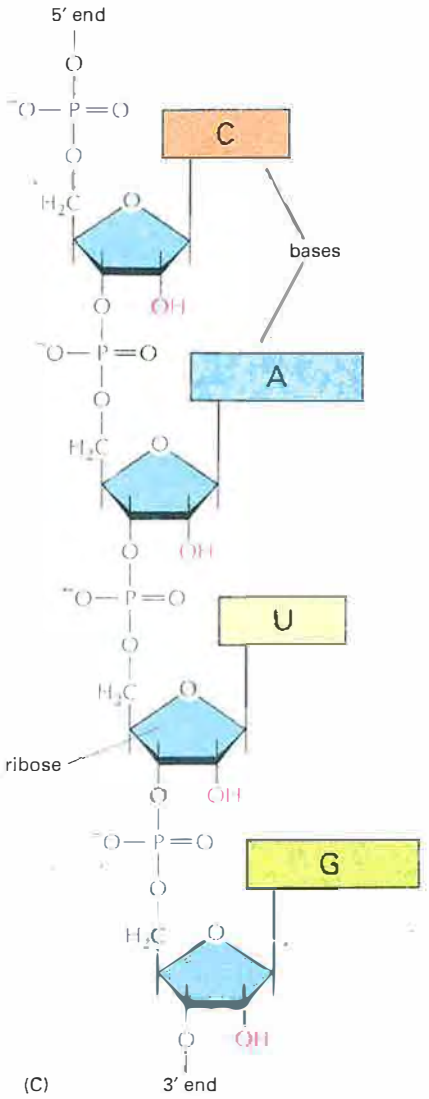


Figure 6–4 The chemical structure of RNA. (A) RNA contains the sugar ribose, which differs from deoxyribose, the sugar used in DNA, by the presence of an additional –OH group. (B) RNA contains the base uracil, which differs from thymine, the equivalent base in DNA, by the absence of a $-CH_3$ group. (C) A short length of RNA. The phosphodiester chemical linkage between nucleotides in RNA is the same as that in DNA.



Transcription begins with the opening and unwinding of a small portion of the DNA double helix to expose the bases on each DNA strand. One of the two strands of the DNA double helix then acts as a template for the synthesis of an RNA molecule. As in DNA replication, the nucleotide sequence of the RNA chain is determined by the complementary base-pairing between incoming nucleotides and the DNA template. When a good match is made, the incoming ribonucleotide is covalently linked to the growing RNA chain in an enzymatically catalyzed reaction. The RNA chain produced by transcription—the *transcript*—is therefore elongated one nucleotide at a time, and it has a nucleotide sequence that is exactly complementary to the strand of DNA used as the template (Figure 6–7).

Transcription, however, differs from DNA replication in several crucial ways. Unlike a newly formed DNA strand, the RNA strand does not remain hydrogen-bonded to the DNA template strand. Instead, just behind the region where the ribonucleotides are being added, the RNA chain is displaced and the DNA helix re-forms. Thus, the RNA molecules produced by transcription are released from the DNA template as single strands. In addition, because they are copied from only a limited region of the DNA, RNA molecules are much shorter than DNA molecules. A DNA molecule in a human chromosome can be up to 250 million nucleotide-pairs long; in contrast, most RNAs are no more than a few thousand nucleotides long, and many are considerably shorter.

The enzymes that perform transcription are called **RNA polymerases**. Like the DNA polymerase that catalyzes DNA replication (discussed in Chapter 5), RNA polymerases catalyze the formation of the phosphodiester bonds that link the nucleotides together to form a linear chain. The RNA polymerase moves stepwise along the DNA, unwinding the DNA helix just ahead of the active site for polymerization to expose a new region of the template strand for complementary base-pairing. In this way, the growing RNA chain is extended by one nucleotide at a time in the 5′-to-3′ direction (Figure 6–8). The substrates are nucleoside triphosphates (ATP, CTP, UTP, and GTP); as for DNA replication, a hydrolysis of high-energy bonds provides the energy needed to drive the reaction forward (see Figure 5–4).

The almost immediate release of the RNA strand from the DNA as it is synthesized means that many RNA copies can be made from the same gene in a

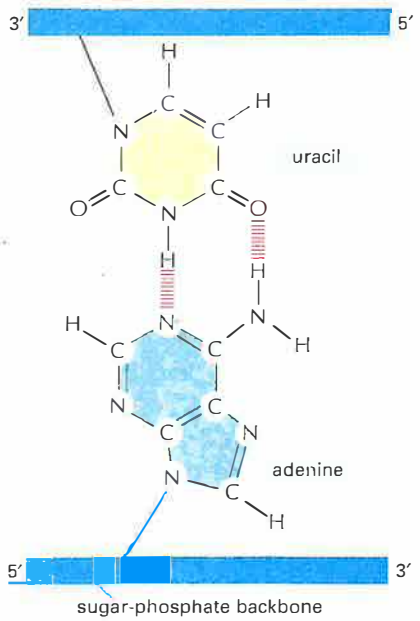


Figure 6–5 Uracil forms base pairs with adenine. The absence of a methyl group in U has no effect on base-pairing; thus, U–A base pairs closely resemble T–A base pairs (see Figure 4–4).

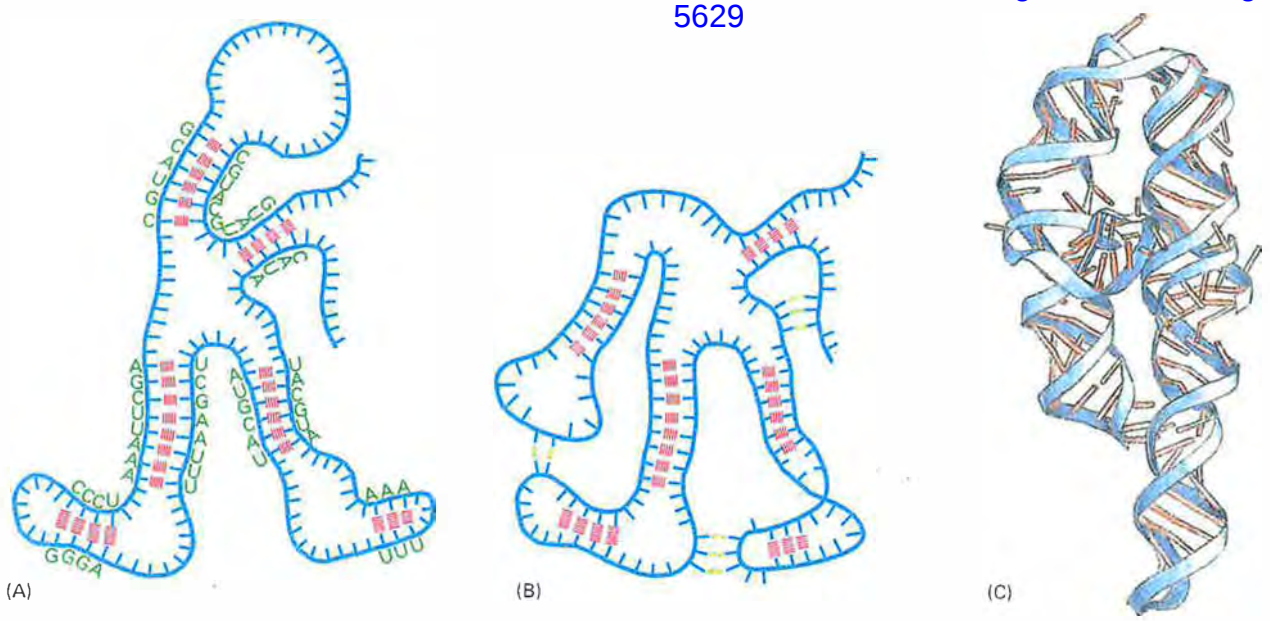


Figure 6–6 RNA can fold into specific structures. RNA is largely single-stranded, but it often contains short stretches of nucleotides that can form conventional base-pairs with complementary sequences found elsewhere on the same molecule. These interactions, along with additional "nonconventional" base-pair interactions, allow an RNA molecule to fold into a three-dimensional structure that is determined by its sequence of nucleotides. (A) Diagram of a folded RNA structure showing only conventional base-pair interactions; (B) structure with both conventional *(red)* and nonconventional *(green)* base-pair interactions; (C) structure of an actual RNA, a portion of a group I intron (see Figure 6–36). Each conventional base-pair interaction is indicated by a "rung" in the double helix. Bases in other configurations are indicated by broken rungs.

relatively short time, the synthesis of additional RNA molecules being started before the first RNA is completed (Figure 6–9). When RNA polymerase molecules follow hard on each other's heels in this way, each moving at about 20 nucleotides per second (the speed in eucaryotes), over a thousand transcripts can be synthesized in an hour from a single gene.

Although RNA polymerase catalyzes essentially the same chemical reaction as DNA polymerase, there are some important differences between the two enzymes. First, and most obvious, RNA polymerase catalyzes the linkage of ribonucleotides, not deoxyribonucleotides. Second, unlike the DNA polymerases involved in DNA replication, RNA polymerases can start an RNA chain without a primer. This difference may exist because transcription need not be as accurate as DNA replication (see Table 5–1, p. 243). Unlike DNA, RNA does not permanently store genetic information in cells. RNA polymerases make about one mistake for every 10^4 nucleotides copied into RNA (compared with an error rate for direct copying by DNA polymerase of about one in 10^7 nucleotides), and the consequences of an error in RNA transcription are much less significant than that in DNA replication.

Although RNA polymerases are not nearly as accurate as the DNA polymerases that replicate DNA, they nonetheless have a modest proofreading mechanism. If the incorrect ribonucleotide is added to the growing RNA chain, the polymerase can back up, and the active site of the enzyme can perform an excision reaction that mimics the reverse of the polymerization reaction, except that water instead of pyrophosphate is used (see Figure 5–4). RNA polymerase hovers around a misincorporated ribonucleotide longer than it does for a correct addition, causing excision to be favored for incorrect nucleotides. However, RNA polymerase also excises many correct bases as part of the cost for improved accuracy.

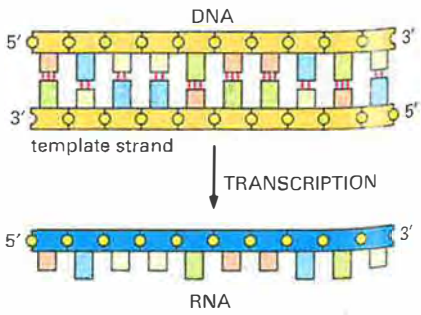


Figure 6–7 DNA transcription produces a single-stranded RNA molecule that is complementary to one strand of DNA.

Cells Produce Several Types of RNA

The majority of genes carried in a cell's DNA specify the amino acid sequence of proteins; the RNA molecules that are copied from these genes (which ultimately direct the synthesis of proteins) are called **messenger RNA (mRNA)** molecules.

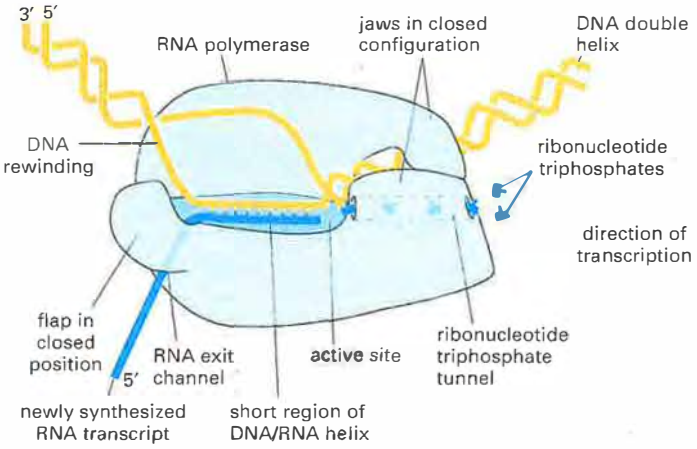


Figure 6–8 DNA is transcribed by the enzyme RNA polymerase. The RNA polymerase *(pale blue)* moves stepwise along the DNA, unwinding the DNA helix at its active site. As it progresses, the polymerase adds nucleotides (here, *small "T" shapes)* one by one to the RNA chain at the polymerization site using an exposed DNA strand as a template. The RNA transcript is thus a single-stranded complementary copy of one of the two DNA strands. The polymerase has a rudder (see Figure 6–11) that displaces the newly formed RNA, allowing the two strands of DNA behind the polymerase to rewind. A short region of DNA/RNA helix (approximately nine nucleotides in length) is therefore formed only transiently, and a "window" of DNA/RNA helix therefore moves along the DNA with the polymerase. The incoming nucleotides are in the form of ribonucleoside triphosphates (ATP, UTP, CTP, and GTP), and the energy stored in their phosphate–phosphate bonds provides the driving force for the polymerization reaction (see Figure 5–4). (Adapted from a figure kindly supplied by Robert Landick.)

The final product of a minority of genes, however, is the RNA itself. Careful analysis of the complete DNA sequence of the genome of the yeast *S. cerevisiae* has uncovered well over 750 genes (somewhat more than 10% of the total number of yeast genes) that produce RNA as their final product, although this number includes multiple copies of some highly repeated genes. These RNAs, like proteins, serve as enzymatic and structural components for a wide variety of processes in the cell. In Chapter 5 we encountered one of those RNAs, the template carried by the enzyme telomerase. Although not all of their functions are known, we see in this chapter that some *small nuclear RNA* (*snRNA*) molecules direct the splicing of pre-mRNA to form mRNA, that *ribosomal RNA (rRNA)* molecules form the core of ribosomes, and that *transfer RNA (tRNA)* molecules form the adaptors that select amino acids and hold them in place on a ribosome for incorporation into protein (Table 6–1).

Each transcribed segment of DNA is called a *transcription unit*. In eucaryotes, a transcription unit typically carries the information of just one gene, and therefore codes for either a single RNA molecule or a single protein (or group of related proteins if the initial RNA transcript is spliced in more than one way to produce different mRNAs). In bacteria, a set of adjacent genes is often transcribed as a unit; the resulting mRNA molecule therefore carries the information for several distinct proteins.

Overall, RNA makes up a few percent of a cell's dry weight. Most of the RNA in cells is rRNA; mRNA comprises only 3–5% of the total RNA in a typical mammalian cell. The mRNA population is made up of tens of thousands of different species, and there are on average only 10–15 molecules of each species of mRNA present in each cell.

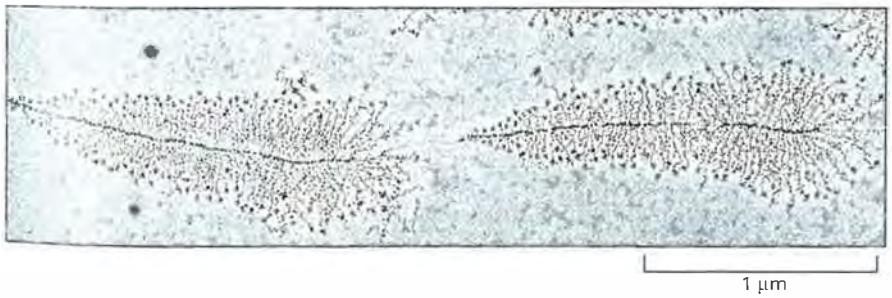


Figure 6–9 Transcription of two genes as observed under the electron microscope. The micrograph shows many molecules of RNA polymerase simultaneously transcribing each of two adjacent genes. Molecules of RNA polymerase are visible as a series of dots along the DNA with the newly synthesized transcripts (fine threads) attached to them. The RNA molecules (ribosomal RNAs) shown in this example are not translated into protein but are instead used directly as components of ribosomes, the machines on which translation takes place. The particles at the 5′ end (the free end) of each rRNA transcript are believed to reflect the beginnings of ribosome assembly. From the lengths of the newly synthesized transcripts, it can be deduced that the RNA polymerase molecules are transcribing from left to right. (Courtesy of Ulrich Scheer.)

TABLE 6–1 Principal Types of RNAs Produced in Cells

TYPE OF RNA	FUNCTION
mRNAs	messenger RNAs, code for proteins
rRNAs	ribosomal RNAs, form the basic structure of the ribosome and catalyze protein synthesis
tRNAs	transfer RNAs, central to protein synthesis as adaptors between mRNA and amino acids
snRNAs	small nuclear RNAs, function in a variety of nuclear processes, including the splicing of pre-mRNA
snoRNAs	small nucleolar RNAs, used to process and chemically modify rRNAs
Other noncoding RNAs	function in diverse cellular processes, including telomere synthesis, X-chromosome inactivation, and the transport of proteins into the ER

Signals Encoded in DNA Tell RNA Polymerase Where to Start and Stop

To transcribe a gene accurately, RNA polymerase must recognize where on the genome to start and where to finish. The way in which RNA polymerases perform these tasks differs somewhat between bacteria and eucaryotes. Because the process in bacteria is simpler, we look there first.

The initiation of transcription is an especially important step in gene expression because it is the main point at which the cell regulates which proteins are to be produced and at what rate. Bacterial RNA polymerase is a multisubunit complex. A detachable subunit, called *sigma (σ) factor*, is largely responsible for its ability to read the signals in the DNA that tell it where to begin transcribing (Figure 6–10). RNA polymerase molecules adhere only weakly to the bacterial DNA when they collide with it, and a polymerase molecule typically slides rapidly along the long DNA molecule until it dissociates again. However, when the polymerase slides into a region on the DNA double helix called a **promoter**, a special sequence of nucleotides indicating the starting point for RNA synthesis, it binds tightly to it. The polymerase, using its σ factor, recognizes this DNA sequence by making specific contacts with the portions of the bases that are exposed on the outside of the helix *(Step 1* in Figure 6–10).

After the RNA polymerase binds tightly to the promoter DNA in this way, it opens up the double helix to expose a short stretch of nucleotides on each strand *(Step 2* in Figure 6–10). Unlike a DNA helicase reaction (see Figure 5–15), this limited opening of the helix does not require the energy of ATP hydrolysis. Instead, the polymerase and DNA both undergo reversible structural changes that result in a more energetically favorable state. With the DNA unwound, one of the two exposed DNA strands acts as a template for complementary base-pairing with incoming ribonucleotides (see Figure 6–7), two of which are joined together by the polymerase to begin an RNA chain. After the first ten or so nucleotides of RNA have been synthesized (a relatively inefficient process during which polymerase synthesizes and discards short nucleotide oligomers), the σ factor relaxes its tight hold on the polymerase and evenutally dissociates from it. During this process, the polymerase undergoes additional structural changes that enable it to move forward rapidly, transcribing without the σ factor *(Step 4* in Figure 6–10). Chain elongation continues (at a speed of approximately 50 nucleotides/sec for bacterial RNA polymerases) until the enzyme encounters a second signal in the DNA, the **terminator** (described below), where the polymerase halts and releases both the DNA template and the newly made RNA chain *(Step 7* in Figure 6–10). After the polymerase has been released at a terminator, it reassociates with a free σ factor and searches for a new promoter, where it can begin the process of transcription again.

Several structural features of bacterial RNA polymerase make it particularly adept at performing the transcription cycle just described. Once the σ factor

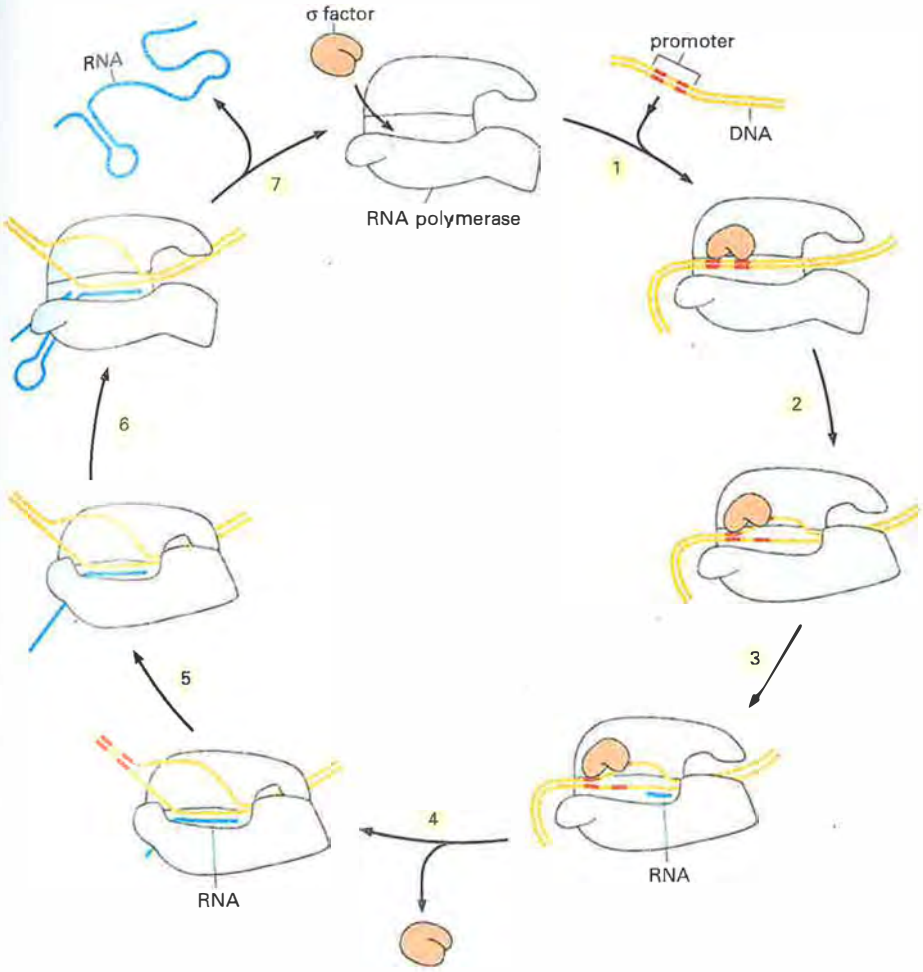


Figure 6–10 The transcription cycle of bacterial RNA polymerase. In step 1, the RNA polymerase holoenzyme (core polymerase plus σ factor) forms and then locates a promoter (see Figure 6–12). The polymerase unwinds the DNA at the position at which transcription is to begin (step 2) and begins transcribing (step 3). This initial RNA synthesis (sometimes called "abortive initiation") is relatively inefficient. However, once RNA polymerase has managed to synthesize about 10 nucleotides of RNA, σ relaxes its grip, and the polymerase undergoes a series of conformational changes (which probably includes a tightening of its jaws and the placement of RNA in the exit channel [see Figure 6–11]). The polymerase now shifts to the elongation mode of RNA synthesis (step 4), moving rightwards along the DNA in this diagram. During the elongation mode (step 5) transcription is highly processive, with the polymerase leaving the DNA template and releasing the newly transcribed RNA only when it encounters a termination signal (step 6). Termination signals are encoded in DNA and many function by forming an RNA structure that destabilizes the polymerase's hold on the RNA, as shown here. In bacteria, all RNA molecules are synthesized by a single type of RNA polymerase and the cycle depicted in the figure therefore applies to the production of mRNAs as well as structural and catalytic RNAs. (Adapted from a figure kindly supplied by Robert Landick.)

positions the polymerase on the promoter and the template DNA has been unwound and pushed to the active site, a pair of moveable jaws is thought to clamp onto the DNA (Figure 6–11). When the first 10 nucleotides have been transcribed, the dissociation of σ allows a flap at the back of the polymerase to

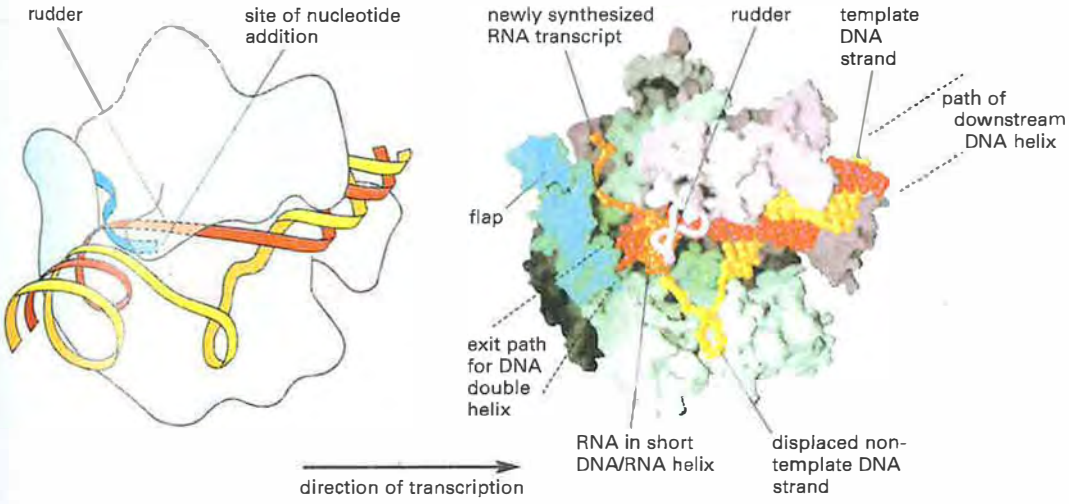
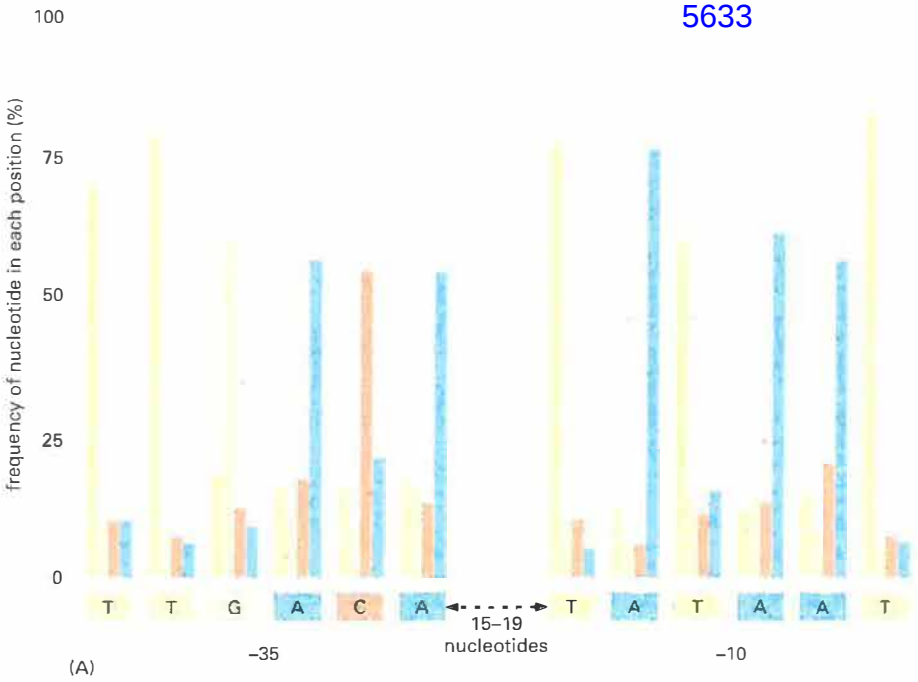


Figure 6–11 The structure of a bacterial RNA polymerase. Two depictions of the three-dimensional structure of a bacterial RNA polymerase, with the DNA and RNA modeled in. This RNA polymerase is formed from four different subunits, indicated by different colors *(right)*. The DNA strand used as a template is red, and the non-template strand is *yellow*. The rudder wedges apart the DNA–RNA hybrid as the polymerase moves. For simplicity only the polypeptide backbone of the rudder is shown in the right-hand figure, and the DNA exiting from the polymerase has been omitted. Because the RNA polymerase is depicted in the elongation mode, the σ factor is absent. (Courtesy of Seth Darst.)



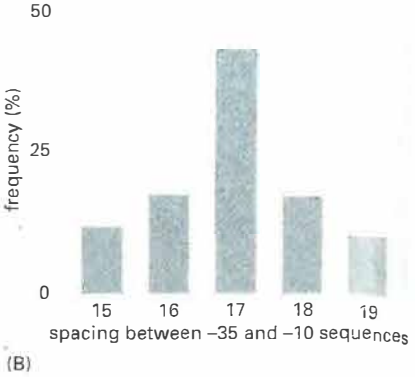


Figure 6–12 Consensus sequence for the major class of *E. coli* promoters. (A) The promoters are characterized by two hexameric DNA sequences, the –35 sequence and the –10 sequence named for their approximate location relative to the start point of transcription (designated +1). For convenience, the nucleotide sequence of a single strand of DNA is shown; in reality the RNA polymerase recognizes the promoter as double-stranded DNA. On the basis of a comparison of 300 promoters, the frequencies of the four nucleotides at each position in the –35 and –10 hexamers are given. The consensus sequence, shown *below* the graph, reflects the most common nucleotide found at each position in the collection of promoters. The sequence of nucleotides between the –35 and –10 hexamers shows no significant similarities among promoters. (B) The distribution of spacing between the –35 and –10 hexamers found in *E. coli* promoters. The information displayed in these two graphs applies to *E. coli* promoters that are recognized by RNA polymerase and the major σ factor (designated σ^{70}). As we shall see in the next chapter, bacteria also contain minor σ factors, each of which recognizes a different promoter sequence. Some particularly strong promoters recognized by RNA polymerase and σ^{70} have an additional sequence, located upstream (to the *left*, in the figure) of the –35 hexamer, which is recognized by another subunit of RNA polymerase.

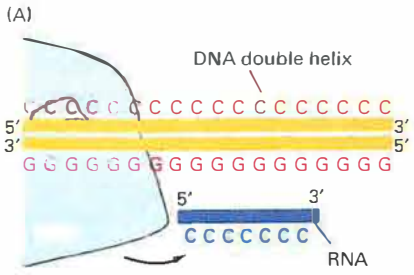
close to form an exit tunnel through which the newly made RNA leaves the enzyme. With the polymerase now functioning in its elongation mode, a rudder-like structure in the enzyme continuously pries apart the DNA–RNA hybrid formed. We can view the series of conformational changes that takes place during transcription initiation as a successive tightening of the enzyme around the DNA and RNA to ensure that it does not dissociate before it has finished transcribing a gene. If an RNA polymerase does dissociate prematurely, it cannot resume synthesis but must start over again at the promoter.

How do the signals in the DNA (termination signals) stop the elongating polymerase? For most bacterial genes a termination signal consists of a string of A–T nucleotide pairs preceded by a two-fold symmetric DNA sequence, which, when transcribed into RNA, folds into a "hairpin" structure through Watson–Crick base-pairing (see Figure 6–10). As the polymerase transcribes across a terminator, the hairpin may help to wedge open the movable flap on the RNA polymerase and release the RNA transcript from the exit tunnel. At the same time, the DNA–RNA hybrid in the active site, which is held together predominantly by U–A base pairs (which are less stable than G–C base pairs because they form two rather than three hydrogen bonds per base pair), is not sufficiently strong enough to hold the RNA in place, and it dissociates causing the release of the polymerase from the DNA, perhaps by forcing open its jaws. Thus, in some respects, transcription termination seems to involve a reversal of the structural transitions that happen during initiation. The process of termination also is an example of a common theme in this chapter: the ability of RNA to fold into specific structures figures prominantly in many aspects of decoding the genome.

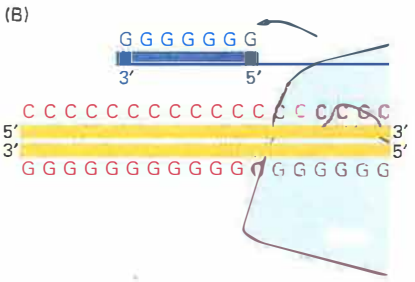
Transcription Start and Stop Signals Are Heterogeneous in Nucleotide Sequence

As we have just seen, the processes of transcription initiation and termination involve a complicated series of structural transitions in protein, DNA, and RNA molecules. It is perhaps not surprising that the signals encoded in DNA that specify these transitions are difficult for researchers to recognize. Indeed, a comparison of many different bacterial promoters reveals that they are heterogeneous in DNA sequence. Nevertheless, they all contain related sequences, reflecting in part aspects of the DNA that are recognized directly by the σ factor. These common features are often summarized in the form of a *consensus sequence* (Figure 6–12). In general, a consensus nucleotide sequence is derived

Figure 6–13 The importance of RNA polymerase orientation. The DNA strand serving as template must be traversed in a 3′ to 5′ direction, as illustrated in Figure 6–9. Thus, the direction of RNA polymerase movement determines which of the two DNA strands is to serve as a template for the synthesis of RNA, as shown in (A) and (B). Polymerase direction is, in turn, determined by the orientation of the promoter sequence, the site at which the RNA polymerase begins transcription.



an RNA polymerase that moves from left to right makes RNA by using the bottom strand as a template



an RNA polymerase that moves from right to left makes RNA by using the top strand as a template

by comparing many sequences with the same basic function and tallying up the most common nucleotide found at each position. It therefore serves as a summary or "average" of a large number of individual nucleotide sequences.

One reason that individual bacterial promoters differ in DNA sequence is that the precise sequence determines the strength (or number of initiation events per unit time) of the promoter. Evolutionary processes have thus fine-tuned each promoter to initiate as often as necessary and have created a wide spectrum of promoters. Promoters for genes that code for abundant proteins are much stronger than those associated with genes that encode rare proteins, and their nucleotide sequences are responsible for these differences.

Like bacterial promoters, transcription terminators also include a wide range of sequences, with the potential to form a simple RNA structure being the most important common feature. Since an almost unlimited number of nucleotide sequences have this potential, terminator sequences are much more heterogeneous than those of promoters.

We have discussed bacterial promoters and terminators in some detail to illustrate an important point regarding the analysis of genome sequences. Although we know a great deal about bacterial promoters and terminators and can develop consensus sequences that summarize their most salient features, their variation in nucleotide sequence makes it difficult for researchers (even when aided by powerful computers) to definitively locate them simply by inspection of the nucleotide sequence of a genome. When we encounter analogous types of sequences in eucaryotes, the problem of locating them is even more difficult. Often, additional information, some of it from direct experimentation, is needed to accurately locate the short DNA signals contained in genomes.

Promoter sequences are asymmetric (see Figure 6–12), and this feature has important consequences for their arrangement in genomes. Since DNA is double-stranded, two different RNA molecules could in principle be transcribed from any gene, using each of the two DNA strands as a template. However a gene typically has only a single promoter, and because the nucleotide sequences of bacterial (as well as eucaryotic) promoters are asymmetric the polymerase can bind in only one orientation. The polymerase thus has no option but to transcribe the one DNA strand, since it can synthesize RNA only in the 5′ to 3′ direction (Figure 6–13). The choice of template strand for each gene is therefore determined by the location and orientation of the promoter. Genome sequences reveal that the DNA strand used as the template for RNA synthesis varies from gene to gene (Figure 6–14; see also Figure 1–31).

Having considered transcription in bacteria, we now turn to the situation in eucaryotes, where the synthesis of RNA molecules is a much more elaborate affair.

Transcription Initiation in Eucaryotes Requires Many Proteins

In contrast to bacteria, which contain a single type of RNA polymerase, eucaryotic nuclei have three, called *RNA polymerase I, RNA polymerase II,* and *RNA*

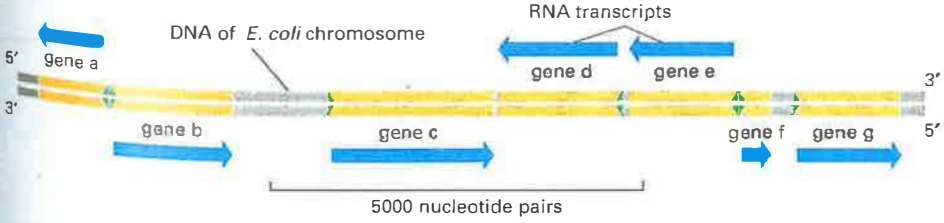


Figure 6–14 Directions of transcription along a short portion of a bacterial chromosome. Some genes are transcribed using one DNA strand as a template, while others are transcribed using the other DNA strand. The direction of transcription is determined by the promoter at the beginning of each gene *(green arrowheads).* Approximately 0.2% (9000 base pairs) of the *E. coli* chromosome is depicted here. The genes transcribed from *left* to *right* use the bottom DNA strand as the template; those transcribed from *right* to *left* use the top strand as the template.

TABLE 6–2 The Three RNA Polymerases in Eucaryotic Cells

TYPE OF POLYMERASE	GENES TRANSCRIBED
RNA polymerase I	5.8S, 18S, and 28S rRNA genes
RNA polymerase II	all protein-coding genes, plus snoRNA genes and some snRNA genes
RNA polymerase III	tRNA genes, 5S rRNA genes, some snRNA genes and genes for other small RNAs

polymerase III. The three polymerases are structurally similar to one another (and to the bacterial enzyme). They share some common subunits and many structural features, but they transcribe different types of genes (Table 6–2). RNA polymerases I and III transcribe the genes encoding transfer RNA, ribosomal RNA, and various small RNAs. RNA polymerase II transcribes the vast majority of genes, including all those that encode proteins, and our subsequent discussion therefore focuses on this enzyme.

Although eucaryotic RNA polymerase II has many structural similarities to bacterial RNA polymerase (Figure 6–15), there are several important differences in the way in which the bacterial and eucaryotic enzymes function, two of which concern us immediately.

1. While bacterial RNA polymerase (with σ factor as one of its subunits) is able to initiate transcription on a DNA template *in vitro* without the help of additional proteins, eucaryotic RNA polymerases cannot. They require the help of a large set of proteins called *general transcription factors*, which must assemble at the promoter with the polymerase before the polymerase can begin transcription.
2. Eucaryotic transcription initiation must deal with the packing of DNA into nucleosomes and higher order forms of chromatin structure, features absent from bacterial chromosomes.

RNA Polymerase II Requires General Transcription Factors

The discovery that, unlike bacterial RNA polymerase, purified eucaryotic RNA polymerase II could not initiate transcription *in vitro* led to the discovery and purification of the additional factors required for this process. These **general transcription factors** help to position the RNA polymerase correctly at the promoter, aid in pulling apart the two strands of DNA to allow transcription to begin, and release RNA polymerase from the promoter into the elongation mode once transcription has begun. The proteins are "general" because they assemble on all promoters used by RNA polymerase II; consisting of a set of interacting proteins, they are designated as *TFII* (for transcription factor for polymerase II),

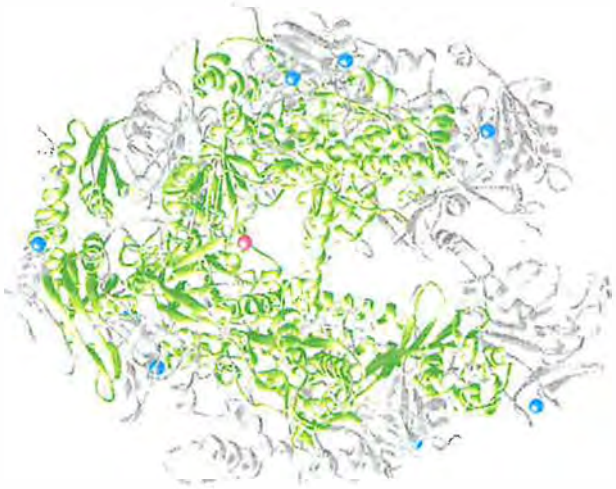


Figure 6–15 Structural similarity between a bacterial RNA polymerase and a eucaryotic RNA polymerase II. Regions of the two RNA polymerases that have similar structures are indicated in *green*. The eucaryotic polymerase is larger than the bacterial enzyme (12 subunits instead of 5), and some of the additional regions are shown in *gray*. The *blue* spheres represent Zn atoms that serve as structural components of the polymerases, and the *red* sphere represents the Mg atom present at the active site, where polymerization takes place. The RNA polymerases in all modern-day cells (bacteria, archaea, and eucaryotes) are closely related, indicating that the basic features of the enzyme were in place before the divergence of the three major branches of life. (Courtesy of P. Cramer and R. Kornberg.)

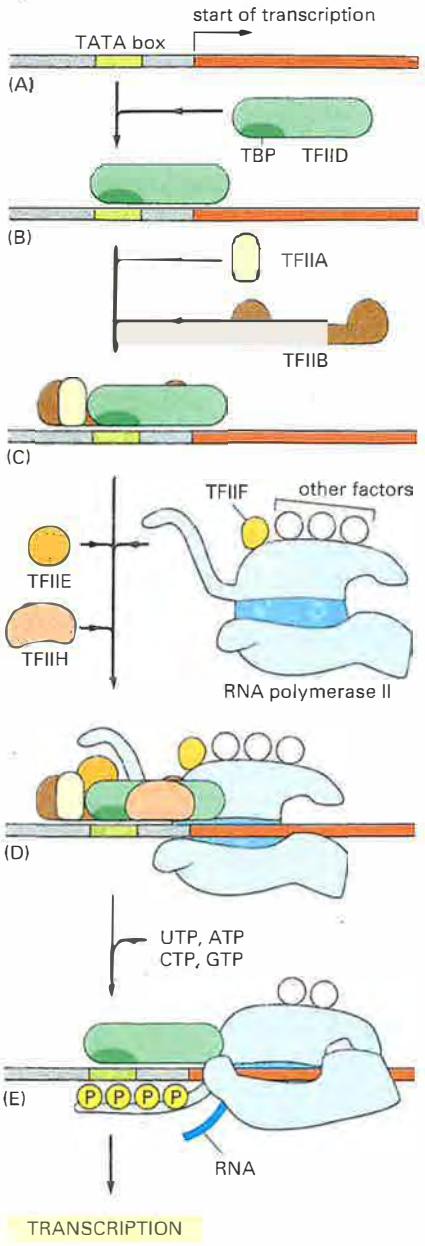
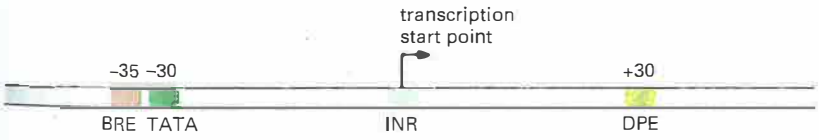


Figure 6–16 Initiation of transcription of a eucaryotic gene by RNA polymerase II. To begin transcription, RNA polymerase requires a number of general transcription factors (called TFIIA, TFIIB, and so on). (A) The promoter contains a DNA sequence called the TATA box, which is located 25 nucleotides away from the site at which transcription is initiated. (B) The TATA box is recognized and bound by transcription factor TFIID, which then enables the adjacent binding of TFIIB (C). For simplicity the DNA distortion produced by the binding of TFIID (see Figure 6–18) is not shown. (D) The rest of the general transcription factors, as well as the RNA polymerase itself, assemble at the promoter. (E) TFIIH then uses ATP to pry apart the DNA double helix at the transcription start point, allowing transcription to begin. TFIIH also phosphorylates RNA polymerase II, changing its conformation so that the polymerase is released from the general factors and can begin the elongation phase of transcription. As shown, the site of phosphorylation is a long C-terminal polypeptide tail that extends from the polymerase molecule. The assembly scheme shown in the figure was deduced from experiments performed *in vitro*, and the exact order in which the general transcription factors assemble on promoters in cells is not known with certainty. In some cases, the general factors are thought to first assemble with the polymerase, with the whole assembly subsequently binding to the DNA in a single step. The general transcription factors have been highly conserved in evolution; some of those from human cells can be replaced in biochemical experiments by the corresponding factors from simple yeasts.

and listed as TFIIA, TFIIB, and so on. In a broad sense, the eucaryotic general transcription factors carry out functions equivalent to those of the σ factor in bacteria.

Figure 6–16 shows how the general transcription factors assemble *in vitro* at promoters used by RNA polymerase II. The assembly process starts with the binding of the general transcription factor TFIID to a short double-helical DNA sequence primarily composed of T and A nucleotides. For this reason, this sequence is known as the TATA sequence, or **TATA box**, and the subunit of TFIID that recognizes it is called TBP (for TATA-binding protein). The TATA box is typically located 25 nucleotides upstream from the transcription start site. It is not the only DNA sequence that signals the start of transcription (Figure 6–17), but



element	consensus sequence	general transcription factor
BRE	G/C G/C G/A C G C C	TFIIB
TATA	T A T A A/T A A/T	TBP
INR	C/T C/T A N T/A C/T C/T	TFIID
DPE	A/G G A/T C G T G	TFIID

Figure 6–17 Consensus sequences found in the vicinity of eucaryotic RNA polymerase II start points. The name given to each consensus sequence *(first column)* and the general transcription factor that recognizes it *(last column)* are indicated. N indicates any nucleotide, and two nucleotides separated by a slash indicate an equal probability of either nucleotide at the indicated position. In reality, each consensus sequence is a shorthand representation of a histogram similar to that of Figure 6–12. For most RNA polymerase II transcription start points, only two or three of the four sequences are present. For example, most polymerase II promoters have a TATA box sequence, and those that do not typically have a "strong" INR sequence. Although most of the DNA sequences that influence transcription initiation are located "upstream" of the transcription start point, a few, such as the DPE shown in the figure, are located in the transcribed region.

for most polymerase II promoters, it is the most important. The binding of TFIID causes a large distortion in the DNA of the TATA box (Figure 6–18). This distortion is thought to serve as a physical landmark for the location of an active promoter in the midst of a very large genome, and it brings DNA sequences on both sides of the distortion together to allow for subsequent protein assembly steps. Other factors are then assembled, along with RNA polymerase II, to form a complete *transcription initiation complex* (see Figure 6–16).

After RNA polymerase II has been guided onto the promoter DNA to form a transcription initiation complex, it must gain access to the template strand at the transcription start point. This step is aided by one of the general transcription factors, TFIIH, which contains a DNA helicase. Next, like the bacterial polymerase, polymerase II remains at the promoter, synthesizing short lengths of RNA until it undergoes a conformational change and is released to begin transcribing a gene. A key step in this release is the addition of phosphate groups to the "tail" of the RNA polymerase (known as the CTD or C-terminal domain). This phosphorylation is also catalyzed by TFIIH, which, in addition to a helicase, contains a protein kinase as one of its subunits (see Figure 6–16, D and E). The polymerase can then disengage from the cluster of general transcription factors, undergoing a series of conformational changes that tighten its interaction with DNA and acquiring new proteins that allow it to transcribe for long distances without dissociating.

Once the polymerase II has begun elongating the RNA transcript, most of the general transcription factors are released from the DNA so that they are available to initiate another round of transcription with a new RNA polymerase molecule. As we see shortly, the phosphorylation of the tail of RNA polymerase II also causes components of the RNA processing machinery to load onto the polymerase and thus be in position to modify the newly transcribed RNA as it emerges from the polymerase.

Polymerase II Also Requires Activator, Mediator, and Chromatin-modifying Proteins

The model for transcription initiation just described was established by studying the action of RNA polymerase II and its general transcription factors on purified DNA templates *in vitro*. However, as discussed in Chapter 4, DNA in eucaryotic cells is packaged into nucleosomes, which are further arranged in higher-order chromatin structures. As a result, transcription initiation in a eucaryotic cell is more complex and requires more proteins than it does on purified DNA. First, gene regulatory proteins known as *transcriptional activators*

Figure 6–18 Three-dimensional structure of TBP (TATA-binding protein) bound to DNA. The TBP is the subunit of the general transcription factor TFIID that is responsible for recognizing and binding to the TATA box sequence in the DNA *(red)*. The unique DNA bending caused by TBP—two kinks in the double helix separated by partly unwound DNA—may serve as a landmark that helps to attract the other general transcription factors. TBP is a single polypeptide chain that is folded into two very similar domains *(blue* and *green)*. (Adapted from J.L. Kim et al., *Nature* 365:520–527, 1993.)

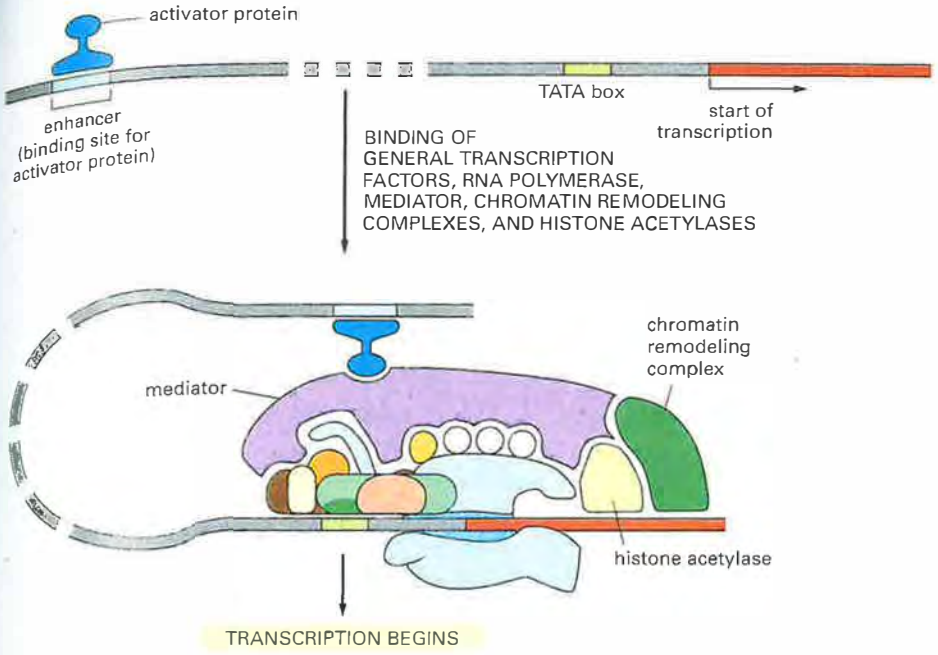


Figure 6–19 Transcription initiation by RNA polymerase II in a eucaryotic cell. Transcription initiation *in vivo* requires the presence of transcriptional activator proteins. As described in Chapter 7, these proteins bind to specific short sequences in DNA. Although only one is shown here, a typical eucaryotic gene has many activator proteins, which together determine its rate and pattern of transcription. Sometimes acting from a distance of several thousand nucleotide pairs (indicated by the dashed DNA molecule), these gene regulatory proteins help RNA polymerase, the general factors, and the mediator all to assemble at the promoter. In addition, activators attract ATP-dependent chromatin-remodeling complexes and histone acetylases.

As discussed in Chapter 4, the "default" state of chromatin is probably the 30-nm filament, and this is likely to be a form of DNA upon which transcription is initiated. For simplicity, it is not shown in the figure.

bind to specific sequences in DNA and help to attract RNA polymerase II to the start point of transcription (Figure 6–19). This attraction is needed to help the RNA polymerase and the general transcription factors in overcoming the difficulty of binding to DNA that is packaged in chromatin. We discuss the role of activators in Chapter 7, because they represent one of the main ways in which cells regulate expression of their genes. Here we simply note that their presence on DNA is required for transcription initiation in a eucaryotic cell. Second, eucaryotic transcription initiation *in vivo* requires the presence of a protein complex known as the *mediator*, which allows the activator proteins to communicate properly with the polymerase II and with the general transcription factors. Finally, transcription initiation in the cell often requires the local recruitment of chromatin-modifying enzymes, including chromatin remodeling complexes and histone acetylases (see Figure 6–19). As discussed in Chapter 4, both types of enzymes can allow greater accessibility to the DNA present in chromatin, and by doing so, they facilitate the assembly of the transcription initiation machinery onto DNA.

As illustrated in Figure 6–19, many proteins (well over one hundred individual subunits) must assemble at the start point of transcription to initiate transcription in a eucaryotic cell. The order of assembly of these proteins is probably different for different genes and therefore may not follow a prescribed pathway. In fact, some of these different protein assemblies may interact with each other away from the DNA and be brought to DNA as preformed subcomplexes. For example, the mediator, RNA polymerase II, and some of the general transcription factors can bind to each other in the nucleoplasm and be brought to the DNA as a unit. We return to this issue in Chapter 7, where we discuss the many ways eucaryotic cells can regulate the process of transcription initiation.

Transcription Elongation Produces Superhelical Tension in DNA

Once it has initiated transcription, RNA polymerase does not proceed smoothly along a DNA molecule; rather it moves jerkily, pausing at some sequences and rapidly transcribing through others. Elongating RNA polymerases, both bacterial and eucaryotic, are associated with a series of *elongation factors*, proteins that decrease the likelihood that RNA polymerase will dissociate before it reaches the end of a gene. These factors typically associate with RNA polymerase shortly after initiation has occurred and help polymerases to move through the wide

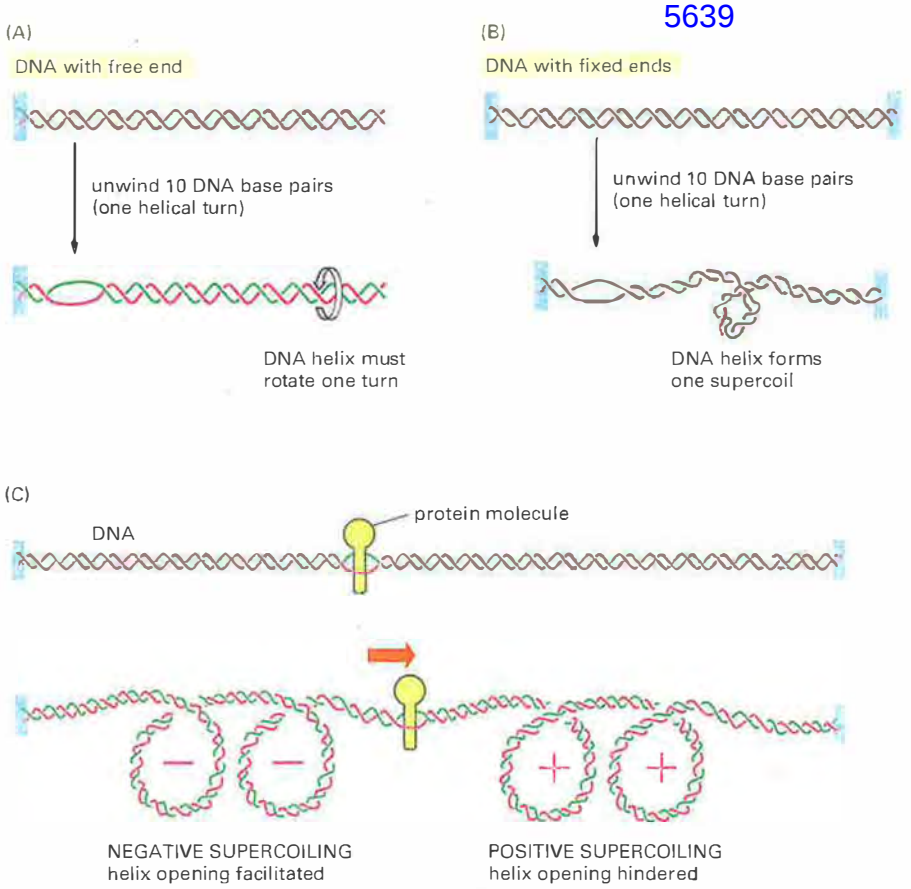


Figure 6–20 Superhelical tension in DNA causes DNA supercoiling. (A) For a DNA molecule with one free end (or a nick in one strand that serves as a swivel), the DNA double helix rotates by one turn for every 10 nucleotide pairs opened. (B) If rotation is prevented, superhelical tension is introduced into the DNA by helix opening. One way of accommodating this tension would be to increase the helical twist from 10 to 11 nucleotide pairs per turn in the double helix that remains in this example; the DNA helix, however, resists such a deformation in a springlike fashion, preferring to relieve the superhelical tension by bending into supercoiled loops. As a result, one DNA supercoil forms in the DNA double helix for every 10 nucleotide pairs opened. The supercoil formed in this case is a positive supercoil. (C) Supercoiling of DNA is induced by a protein tracking through the DNA double helix. The two ends of the DNA shown here are unable to rotate freely relative to each other, and the protein molecule is assumed also to be prevented from rotating freely as it moves. Under these conditions, the movement of the protein causes an excess of helical turns to accumulate in the DNA helix ahead of the protein and a deficit of helical turns to arise in the DNA behind the protein, as shown.

variety of different DNA sequences that are found in genes. Eucaryotic RNA polymerases must also contend with chromatin structure as they move along a DNA template. Experiments have shown that bacterial polymerases, which never encounter nucleosomes *in vivo*, can nonetheless transcribe through them *in vitro*, suggesting that a nucleosome is easily traversed. However, eucaryotic polymerases have to move through forms of chromatin that are more compact than a simple nucleosome. It therefore seems likely that they transcribe with the aid of chromatin remodeling complexes (see pp. 212–213). These complexes may move with the polymerase or may simply seek out and rescue the occasional stalled polymerase. In addition, some elongation factors associated with eucaryotic RNA polymerase facilitate transcription through nucleosomes without requiring additional energy. It is not yet understood how this is accomplished, but these proteins may help to dislodge parts of the nucleosome core as the polymerase transcribes the DNA of a nucleosome.

There is yet another barrier to elongating polymerases, both bacterial and eucaryotic. To discuss this issue, we need first to consider a subtle property inherent in the DNA double helix called **DNA supercoiling**. DNA supercoiling represents a conformation that DNA will adopt in response to superhelical tension; conversely, creating various loops or coils in the helix can create such tension. A simple way of visualizing the topological constraints that cause DNA supercoiling is illustrated in Figure 6–20A. There are approximately 10 nucleotide pairs for every helical turn in a DNA double helix. Imagine a helix whose two ends are fixed with respect to each other (as they are in a DNA circle, such as a bacterial chromosome, or in a tightly clamped loop, as is thought to exist in eucaryotic chromosomes). In this case, one large DNA supercoil will form to compensate for each 10 nucleotide pairs that are opened (unwound). The formation of this supercoil is energetically favorable because it restores a normal helical twist to the base-paired regions that remain, which would otherwise need to be overwound because of the fixed ends.

Superhelical tension is also created as RNA polymerase moves along a stretch of DNA that is anchored at its ends (Figure 6–20C). As long as the polymerase is not free to rotate rapidly (and such rotation is unlikely given the size

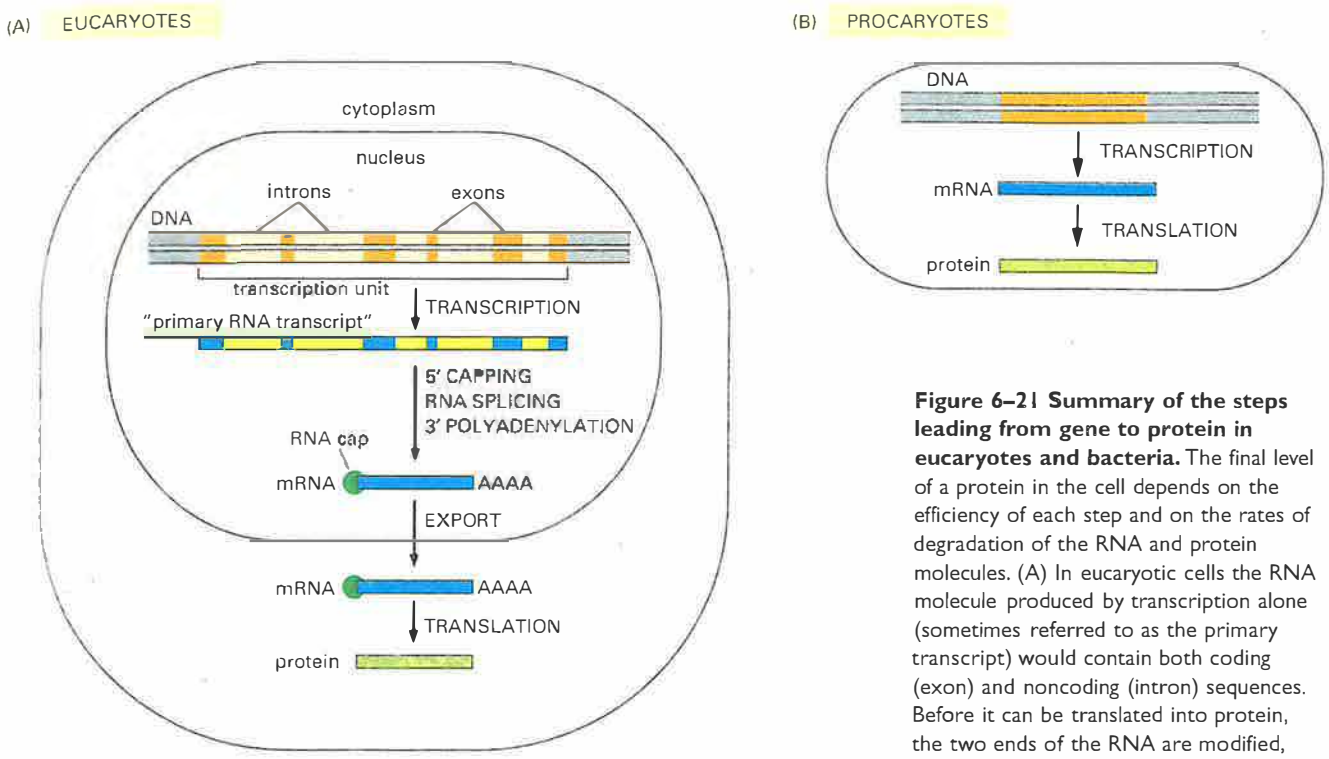


Figure 6–21 Summary of the steps leading from gene to protein in eucaryotes and bacteria. The final level of a protein in the cell depends on the efficiency of each step and on the rates of degradation of the RNA and protein molecules. (A) In eucaryotic cells the RNA molecule produced by transcription alone (sometimes referred to as the primary transcript) would contain both coding (exon) and noncoding (intron) sequences. Before it can be translated into protein, the two ends of the RNA are modified, the introns are removed by an enzymatically catalyzed RNA splicing reaction, and the resulting mRNA is transported from the nucleus to the cytoplasm. Although these steps are depicted as occurring one at a time, in a sequence, in reality they are coupled and different steps can occur simultaneously. For example, the RNA cap is added and splicing typically begins before transcription has been completed. Because of this coupling, complete primary RNA transcripts do not typically exist in the cell. (B) In procaryotes the production of mRNA molecules is much simpler. The 5′ end of an mRNA molecule is produced by the initiation of transcription by RNA polymerase, and the 3′ end is produced by the termination of transcription. Since procaryotic cells lack a nucleus, transcription and translation take place in a common compartment. In fact, translation of a bacterial mRNA often begins before its synthesis has been completed.

of RNA polymerases and their attached transcripts), a moving polymerase generates positive superhelical tension in the DNA in front of it and negative helical tension behind it. For eucaryotes, this situation is thought to provide a bonus: the positive superhelical tension ahead of the polymerase makes the DNA helix more difficult to open, but this tension should facilitate the unwrapping of DNA in nucleosomes, as the release of DNA from the histone core helps to relax positive superhelical tension.

Any protein that propels itself alone along a DNA strand of a double helix tends to generate superhelical tension. In eucaryotes, DNA topoisomerase enzymes rapidly remove this superhelical tension (see p. 251). But, in bacteria, a specialized topoisomerase called *DNA gyrase* uses the energy of ATP hydrolysis to pump supercoils continuously into the DNA, thereby maintaining the DNA under constant tension. These are *negative supercoils,* having the opposite handedness from the *positive supercoils* that form when a region of DNA helix opens (see Figure 6–20B). These negative supercoils are removed from bacterial DNA whenever a region of helix opens, reducing the superhelical tension. DNA gyrase therefore makes the opening of the DNA helix in bacteria energetically favorable compared with helix opening in DNA that is not supercoiled. For this reason, it usually facilitates those genetic processes in bacteria, including the initiation of transcription by bacterial RNA polymerase, that require helix opening (see Figure 6–10).

Transcription Elongation in Eucaryotes Is Tightly Coupled To RNA Processing

We have seen that bacterial mRNAs are synthesized solely by the RNA polymerase starting and stopping at specific spots on the genome. The situation in eucaryotes is substantially different. In particular, transcription is only the first step in a series of reactions that includes the covalent modification of both ends of the RNA and the removal of *intron sequences* that are discarded from the middle of the RNA transcript by the process of *RNA splicing* (Figure 6–21). The modifications of the ends of eucaryotic mRNA are *capping* on the 5′ end and *polyadenylation* of the 3′ end (Figure 6–22). These special ends allow the cell to assess whether both ends of an mRNA molecule are present (and the message is therefore intact) before it exports the RNA sequence from the nucleus for

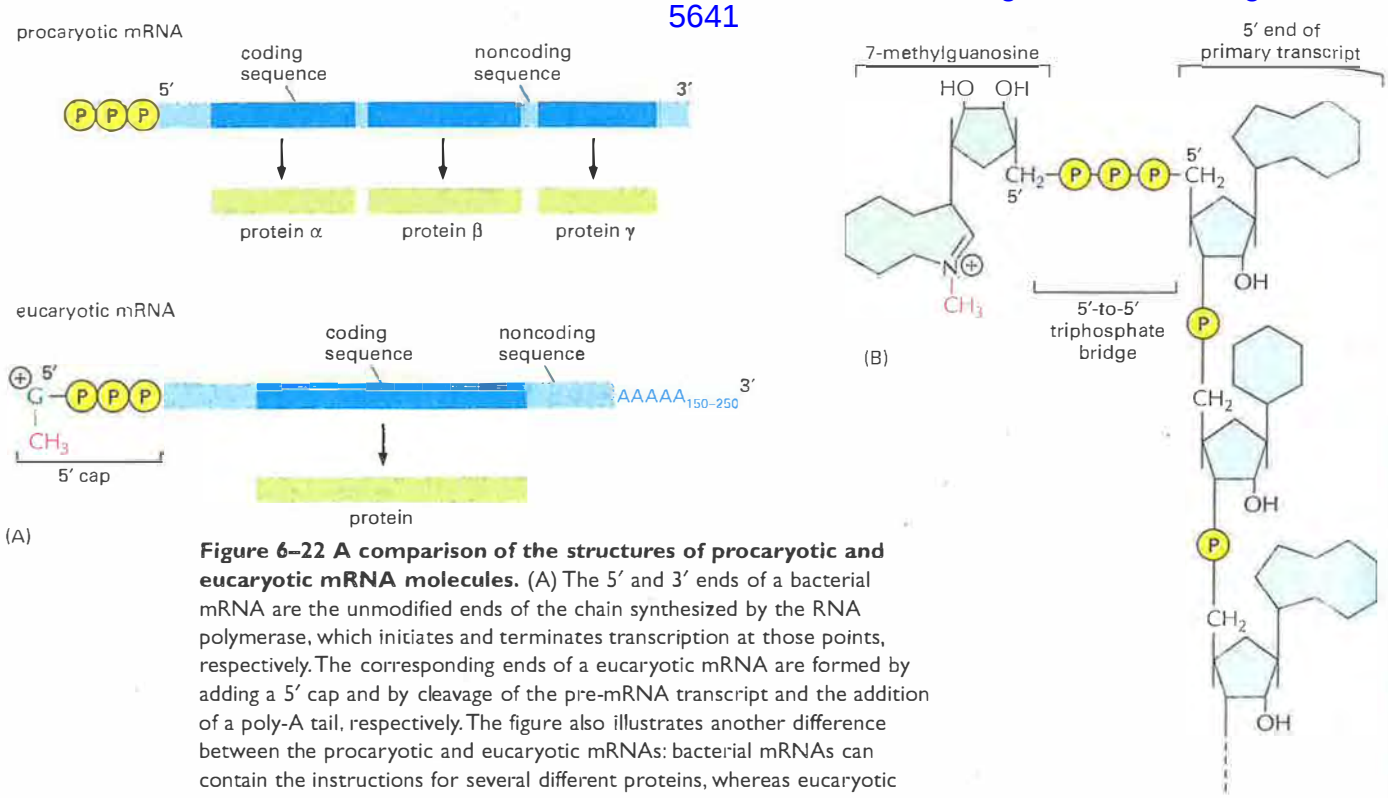


Figure 6–22 A comparison of the structures of procaryotic and eucaryotic mRNA molecules. (A) The 5′ and 3′ ends of a bacterial mRNA are the unmodified ends of the chain synthesized by the RNA polymerase, which initiates and terminates transcription at those points, respectively. The corresponding ends of a eucaryotic mRNA are formed by adding a 5′ cap and by cleavage of the pre-mRNA transcript and the addition of a poly-A tail, respectively. The figure also illustrates another difference between the procaryotic and eucaryotic mRNAs: bacterial mRNAs can contain the instructions for several different proteins, whereas eucaryotic mRNAs nearly always contain the information for only a single protein. (B) The structure of the cap at the 5′ end of eucaryotic mRNA molecules. Note the unusual 5′-to-5′ linkage of the 7-methyl G to the remainder of the RNA. Many eucaryotic mRNAs carry an additional modification: the 2′-hydroxyl group on the second ribose sugar in the mRNA is methylated (not shown).

translation into protein. In Chapter 4, we saw that a typical eucaryotic gene is present in the genome as short blocks of protein-coding sequence (exons) separated by long introns, and RNA splicing is the critically important step in which the different portions of a protein coding sequence are joined together. As we describe next, RNA splicing also provides higher eucaryotes with the ability to synthesize several different proteins from the same gene.

These RNA processing steps are tightly coupled to transcription elongation by an ingenious mechanism. As discussed previously, a key step of the transition of RNA polymerase II to the elongation mode of RNA synthesis is an extensive phosphorylation of the RNA polymerase II tail, called the CTD. This C-terminal domain of the largest subunit consists of a long tandem array of a repeated seven-amino-acid sequence, containing two serines per repeat that can be phosphorylated. Because there are 52 repeats in the CTD of human RNA polymerase II, its complete phosphorylation would add 104 negatively charged phosphate groups to the polymerase. This phosphorylation step not only dissociates the RNA polymerase II from other proteins present at the start point of transcription, it also allows a new set of proteins to associate with the RNA polymerase tail that function in transcription elongation and pre-mRNA processing. As discussed next, some of these processing proteins seem to "hop" from the polymerase tail onto the nascent RNA molecule to begin processing it as it emerges from the RNA polymerase. Thus, RNA polymerase II in its elongation mode can be viewed as an RNA factory that both transcribes DNA into RNA and processes the RNA it produces (Figure 6–23).

RNA Capping Is the First Modification of Eucaryotic Pre-mRNAs

As soon as RNA polymerase II has produced about 25 nucleotides of RNA, the 5′ end of the new RNA molecule is modified by addition of a "cap" that consists of

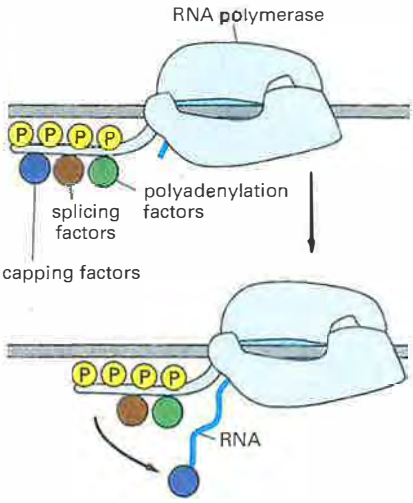


Figure 6–23 The "RNA factory" concept for eucaryotic RNA polymerase II. Not only does the polymerase transcribe DNA into RNA, but it also carries pre-mRNA-processing proteins on its tail, which are then transferred to the nascent RNA at the appropriate time. There are many RNA-processing enzymes, and not all travel with the polymerase. For RNA splicing, for example, only a few critical components are carried on the tail; once transferred to an RNA molecule, they serve as a nucleation site for the remaining components. The RNA-processing proteins first bind to the RNA polymerase tail when it is phosphorylated late in the process of transcription initiation (see Figure 6–16). Once RNA polymerase II finishes transcribing, it is released from DNA, the phosphates on its tail are removed by soluble phosphatases, and it can reinitiate transcription. Only this dephosphorylated form of RNA polymerase II is competent to start RNA synthesis at a promoter.

a modified guanine nucleotide (see Figure 6–22B). The capping reaction is performed by three enzymes acting in succession: one (a phosphatase) removes one phosphate from the 5′ end of the nascent RNA, another (a guanyl transferase) adds a GMP in a reverse linkage (5′ to 5′ instead of 5′ to 3′), and a third (a methyl transferase) adds a methyl group to the guanosine (Figure 6–24). Because all three enzymes bind to the phosphorylated RNA polymerase tail, they are poised to modify the 5′ end of the nascent transcript as soon as it emerges from the polymerase.

The 5′-methyl cap signals the 5′ end of eucaryotic mRNAs, and this landmark helps the cell to distinguish mRNAs from the other types of RNA molecules present in the cell. For example, RNA polymerases I and III produce uncapped RNAs during transcription, in part because these polymerases lack tails. In the nucleus, the cap binds a protein complex called CBC (cap-binding complex), which, as we discuss in subsequent sections, helps the RNA to be properly processed and exported. The 5′ methyl cap also has an important role in the translation of mRNAs in the cytosol as we discuss later in the chapter.

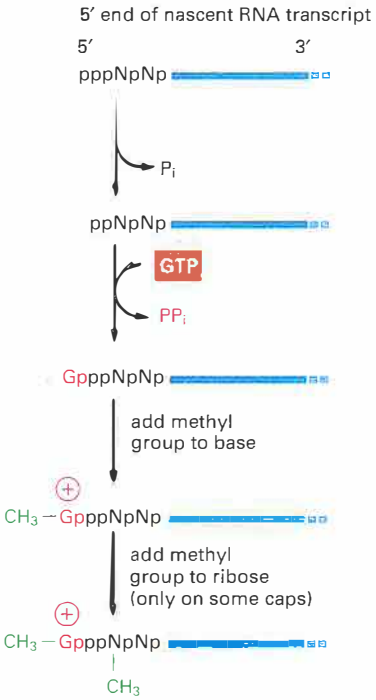


Figure 6–24 The reactions that cap the 5′ end of each RNA molecule synthesized by RNA polymerase II. The final cap contains a novel 5′-to-5′ linkage between the positively charged 7-methyl G residue and the 5′ end of the RNA transcript (see Figure 6–22B). The letter N represents any one of the four ribonucleotides, although the nucleotide that starts an RNA chain is usually a purine (an A or a G). (After A.J. Shatkin, *BioEssays* 7:275–277, 1987. © ICSU Press.)

RNA Splicing Removes Intron Sequences from Newly Transcribed Pre-mRNAs

As discussed in Chapter 4, the protein coding sequences of eucaryotic genes are typically interrupted by noncoding intervening sequences (introns). Discovered in 1977, this feature of eucaryotic genes came as a surprise to scientists, who had been, until that time, familiar only with bacterial genes, which typically consist of a continuous stretch of coding DNA that is directly transcribed into mRNA. In marked contrast, eucaryotic genes were found to be broken up into small pieces of coding **sequence** (*expressed sequences* or **exons**) interspersed with much longer *intervening sequences* or **introns**; thus the coding portion of a eucaryotic gene is often only a small fraction of the length of the gene (Figure 6–25).

Both intron and exon sequences are transcribed into RNA. The intron sequences are removed from the newly synthesized RNA through the process of **RNA splicing**. The vast majority of RNA splicing that takes place in cells functions in the production of mRNA, and our discussion of splicing focuses on this type. It is termed precursor-mRNA (or pre-mRNA) splicing to denote that it occurs on RNA molecules destined to become mRNAs. Only after 5′ and 3′ end processing and splicing have taken place is such RNA termed mRNA.

Each splicing event removes one intron, proceeding through two sequential phosphoryl-transfer reactions known as transesterifications; these join two exons while removing the intron as a "lariat" (Figure 6–26). Since the number of phosphate bonds remains the same, these reactions could in principle take place without nucleoside triphosphate hydrolysis. However, the machinery that catalyzes pre-mRNA splicing is complex, consisting of 5 additional RNA molecules and over 50 proteins, and it hydrolyzes many ATP molecules per splicing event. This complexity is presumably needed to ensure that splicing is highly accurate, while also being sufficiently flexible to deal with the enormous variety of introns found in a typical eucaryotic cell. Frequent mistakes in RNA

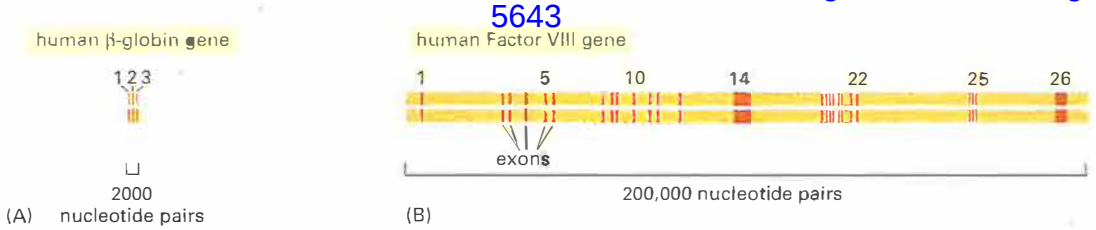


Figure 6–25 Structure of two human genes showing the arrangement of exons and introns. (A) The relatively small β-globin gene, which encodes one of the subunits of the oxygen-carrying protein hemoglobin, contains 3 exons (see also Figure 4–7). (B) The much larger Factor VIII gene contains 26 exons; it codes for a protein (Factor VIII) that functions in the blood-clotting pathway. Mutations in this gene are responsible for the most prevalent form of hemophilia.

splicing would severely harm the cell, as they would result in malfunctioning proteins. We see in Chapter 7 that when rare splicing mistakes do occur, the cell has a "fail-safe" device to eliminate the incorrectly spliced mRNAs.

It may seem wasteful to remove large numbers of introns by RNA splicing. In attempting to explain why it occurs, scientists have pointed out that the exon–intron arrangement would seem to facilitate the emergence of new and useful proteins. Thus, the presence of numerous introns in DNA allows genetic recombination to readily combine the exons of different genes (see p. 462), allowing genes for new proteins to evolve more easily by the combination of parts of preexisting genes. This idea is supported by the observation, described in Chapter 3, that many proteins in present-day cells resemble patchworks composed from a common set of protein pieces, called protein *domains*.

RNA splicing also has a present-day advantage. The transcripts of many eucaryotic genes (estimated at 60% of genes in humans) are spliced in a variety of different ways to produce a set of different mRNAs, thereby allowing a corresponding set of different proteins to be produced from the same gene (Figure 6–27). We discuss additional examples of alternative splicing in Chapter 7, as this is also one of the mechanisms that cells use to change expression of their genes. Rather than being the wasteful process it may have seemed at first sight, RNA splicing enables eucaryotes to increase the already enormous coding potential of their genomes. We shall return to this idea several times in this chapter and the next, but we first need to describe the cellular machinery that performs this remarkable task.

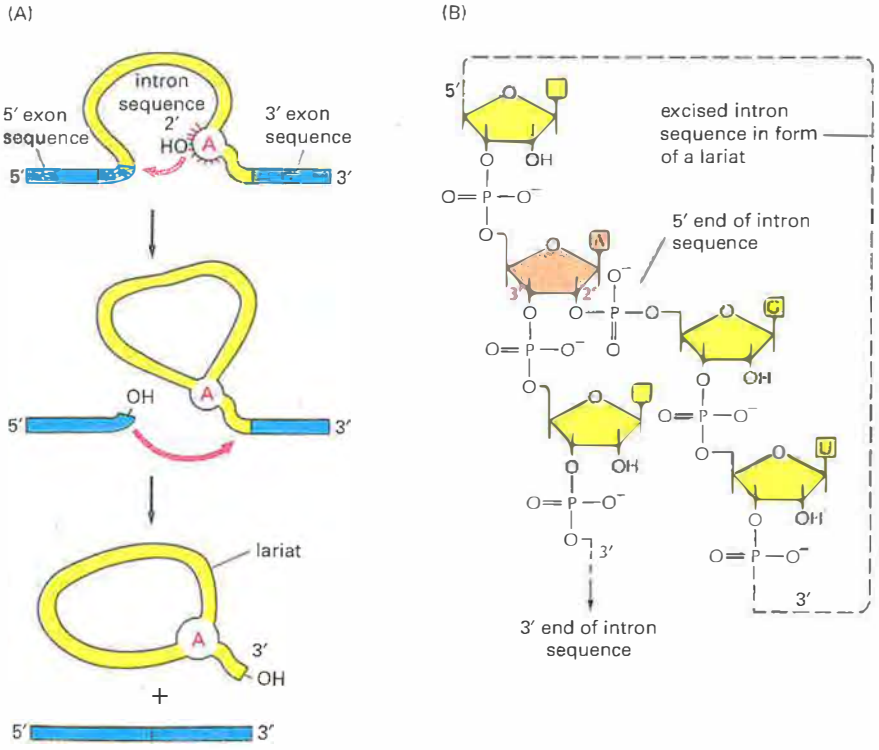


Figure 6–26 The RNA splicing reaction. (A) In the first step, a specific adenine nucleotide in the intron sequence (indicated in *red*) attacks the 5′ splice site and cuts the sugar-phosphate backbone of the RNA at this point. The cut 5′ end of the intron becomes covalently linked to the adenine nucleotide, as shown in detail in (B), thereby creating a loop in the RNA molecule. The released free 3′-OH end of the exon sequence then reacts with the start of the next exon sequence, joining the two exons together and releasing the intron sequence in the shape of a *lariat*. The two exon sequences thereby become joined into a continuous coding sequence; the released intron sequence is degraded in due course.

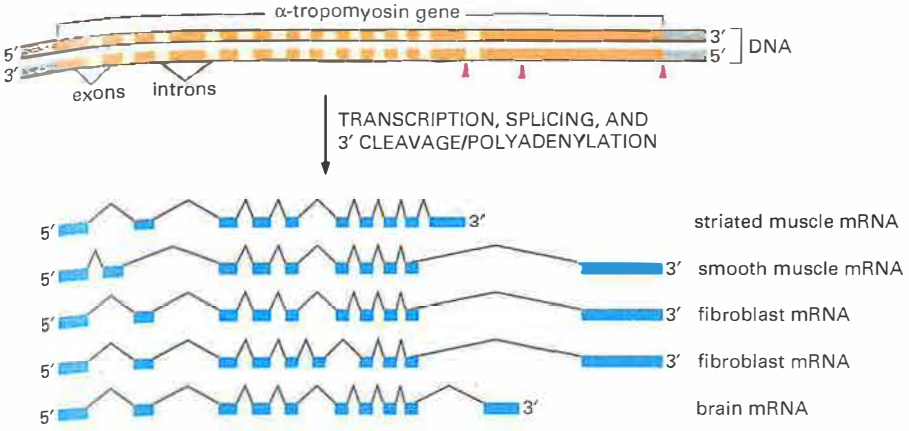


Figure 6–27 Alternative splicing of the α-tropomyosin gene from rat. α-tropomyosin is a coiled-coil protein (see Figure 3–11) that regulates contraction in muscle cells. The primary transcript can be spliced in different ways, as indicated in the figure, to produce distinct mRNAs, which then give rise to variant proteins. Some of the splicing patterns are specific for certain types of cells. For example, the α-tropomyosin made in striated muscle is different from that made from the same gene in smooth muscle. The arrowheads in the top part of the figure demark the sites where cleavage and poly-A addition can occur.

Nucleotide Sequences Signal Where Splicing Occurs

Introns range in size from about 10 nucleotides to over 100,000 nucleotides. Picking out the precise borders of an intron is very difficult for scientists to do (even with the aid of computers) when confronted by a complete genome sequence of a eucaryote. The possibility of alternative splicing compounds the problem of predicting protein sequences solely from a genome sequence. This difficulty constitutes one of the main barriers to identifying all of the genes in a complete genome sequence, and it is the primary reason that we know only the approximate number of genes in, for example, the human genome. Yet each cell in our body recognizes and rapidly excises the appropriate intron sequences with high fidelity. We have seen that intron sequence removal involves three positions on the RNA: the 5′ splice site, the 3′ splice site, and the branch point in the intron sequence that forms the base of the excised lariat. In pre-mRNA splicing, each of these three sites has a consensus nucleotide sequence that is similar from intron to intron, providing the cell with cues on where splicing is to take place (Figure 6–28). However, there is enough variation in each sequence to make it very difficult for scientists to pick out all of the many splicing signals in a genome sequence.

RNA Splicing Is Performed by the Spliceosome

Unlike the other steps of mRNA production we have discussed, RNA splicing is performed largely by RNA molecules instead of proteins. RNA molecules recognize intron–exon borders and participate in the chemistry of splicing. These RNA molecules are relatively short (less than 200 nucleotides each), and there are five of them (U1, U2, U4, U5, and U6) involved in the major form of pre-mRNA splicing. Known as **snRNAs (small nuclear RNAs)**, each is complexed with at least seven protein subunits to form a snRNP (small nuclear ribonucleoprotein). These snRNPs form the core of the **spliceosome**, the large assembly of RNA and protein molecules that performs pre-mRNA splicing in the cell.

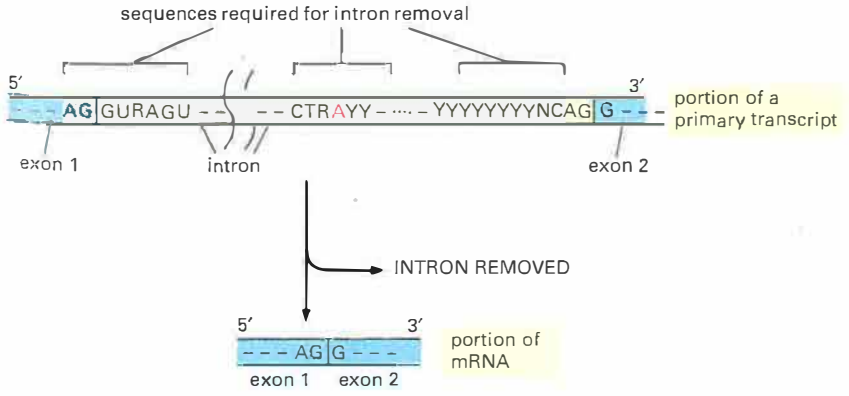


Figure 6–28 The consensus nucleotide sequences in an RNA molecule that signal the beginning and the end of most introns in humans. Only the three blocks of nucleotide sequences shown are required to remove an intron sequence; the rest of the intron can be occupied by any nucleotide. Here A, G, U, and C are the standard RNA nucleotides; R stands for either A or G; Y stands for either C or U. The A highlighted in *red* forms the branch point of the lariat produced by splicing. Only the GU at the start of the intron and the AG at its end are invariant nucleotides in the splicing consensus sequences. The remaining positions (even the branch point A) can be occupied by a variety of nucleotides, although the indicated nucleotides are preferred. The distances along the RNA between the three splicing consensus sequences are highly variable; however, the distance between the branch point and 3′ splice junction is typically much shorter than that between the 5′ splice junction and the branch point.

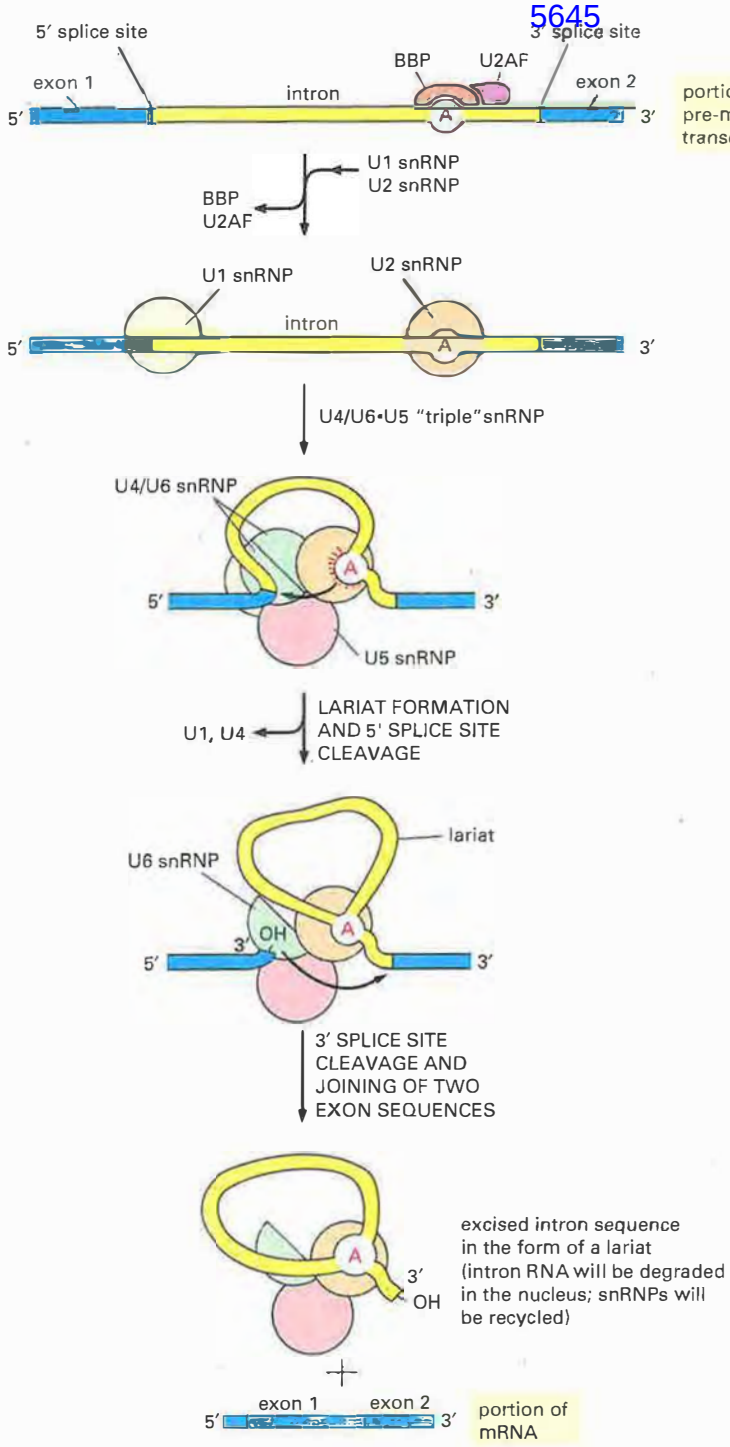


Figure 6–29 The RNA splicing mechanism. RNA splicing is catalyzed by an assembly of snRNPs (shown as *colored circles*) plus other proteins (most of which are not shown), which together constitute the spliceosome. The spliceosome recognizes the splicing signals on a pre-mRNA molecule, brings the two ends of the intron together, and provides the enzymatic activity for the two reaction steps (see Figure 6–26). The branch-point site is first recognized by the BBP (branch-point binding protein) and U2AF, a helper protein. In the next steps, the U2 snRNP displaces BBP and U2AF and forms base pairs with the branch-point site consensus sequence, and the U1 snRNP forms base-pairs with the 5′ splice junction (see Figure 6–30). At this point, the U4/U6•U5 "triple" snRNP enters the spliceosome. In this triple snRNP, the U4 and U6 snRNAs are held firmly together by base-pair interactions and the U5 snRNP is more loosely associated. Several RNA–RNA rearrangements then occur that break apart the U4/U6 base pairs (as shown, the U4 snRNP is ejected from the splicesome before splicing is complete) and allow the U6 snRNP to displace U1 at the 5′ splice junction (see Figure 6–30). Subsequent rearrangements create the active site of the spliceosome and position the appropriate portions of the pre-mRNA substrate for the splicing reaction to occur. Although not shown in the figure, each splicing event requires additional proteins, some of which hydrolyze ATP and promote the RNA–RNA rearrangements.

The spliceosome is a dynamic machine; as we see below, it is assembled on pre-mRNA from separate components, and parts enter and leave it as the splicing reaction proceeds (Figure 6–29). During the splicing reaction, recognition of the 5′ splice junction, the branch point site and the 3′ splice junction is performed largely through base-pairing between the snRNAs and the consensus RNA sequences in the pre-mRNA substrate (Figure 6–30). In the course of splicing, the spliceosome undergoes several shifts in which one set of base-pair interactions is broken and another is formed in its place. For example, U1 is replaced by U6 at the 5′ splice junction (see Figure 6–30A). As we shall see, this type of RNA–RNA rearrangement (in which the formation of one RNA–RNA interaction requires the disruption of another) occurs several times during the splicing reaction. It permits the checking and rechecking of RNA sequences before the chemical reaction is allowed to proceed, thereby increasing the accuracy of splicing.

The Spliceosome Uses ATP Hydrolysis to Produce a Complex Series of RNA–RNA Rearrangements

Although ATP hydrolysis is not required for the chemistry of RNA splicing *per se*, it is required for the stepwise assembly and rearrangements of the spliceosome. Some of the additional proteins that make up the spliceosome are RNA helicases, which use the energy of ATP hydrolysis to break existing RNA–RNA interactions so as to allow the formation of new ones. In fact, all the steps shown previously in Figure 6–29—except the association of BBP with the branch-point site and U1 snRNP with the 5′ splice site—require ATP hydrolysis and additional proteins. In all, more than 50 proteins, including those that form the snRNPs, are required for each splicing event.

The ATP-requiring RNA–RNA rearrangements that take place in the spliceosome occur within the snRNPs themselves and between the snRNPs and the pre-mRNA substrate. One of the most important roles of these rearrangements is the creation of the active catalytic site of the spliceosome. The strategy of creating an active site only after the assembly and rearrangement of splicing components on a pre-mRNA substrate is an important way of preventing wayward splicing.

Perhaps the most surprising feature of the spliceosome is the nature of the catalytic site itself: it is largely (if not exclusively) formed by RNA molecules instead of proteins. In the last section of this chapter we discuss in general terms the structural and chemical properties of RNA that allow it to perform catalysis; here we need only consider that the U2 and U6 snRNAs in the spliceosome form a precise three-dimensional RNA structure that juxtaposes the 5′ splice site of the pre-mRNA with the branch-point site and probably performs the first transesterification reaction (see Figure 6–30C). In a similar way, the 5′ and 3′ splice junctions are brought together (an event requiring the U5 snRNA) to facilitate the second transesterification.

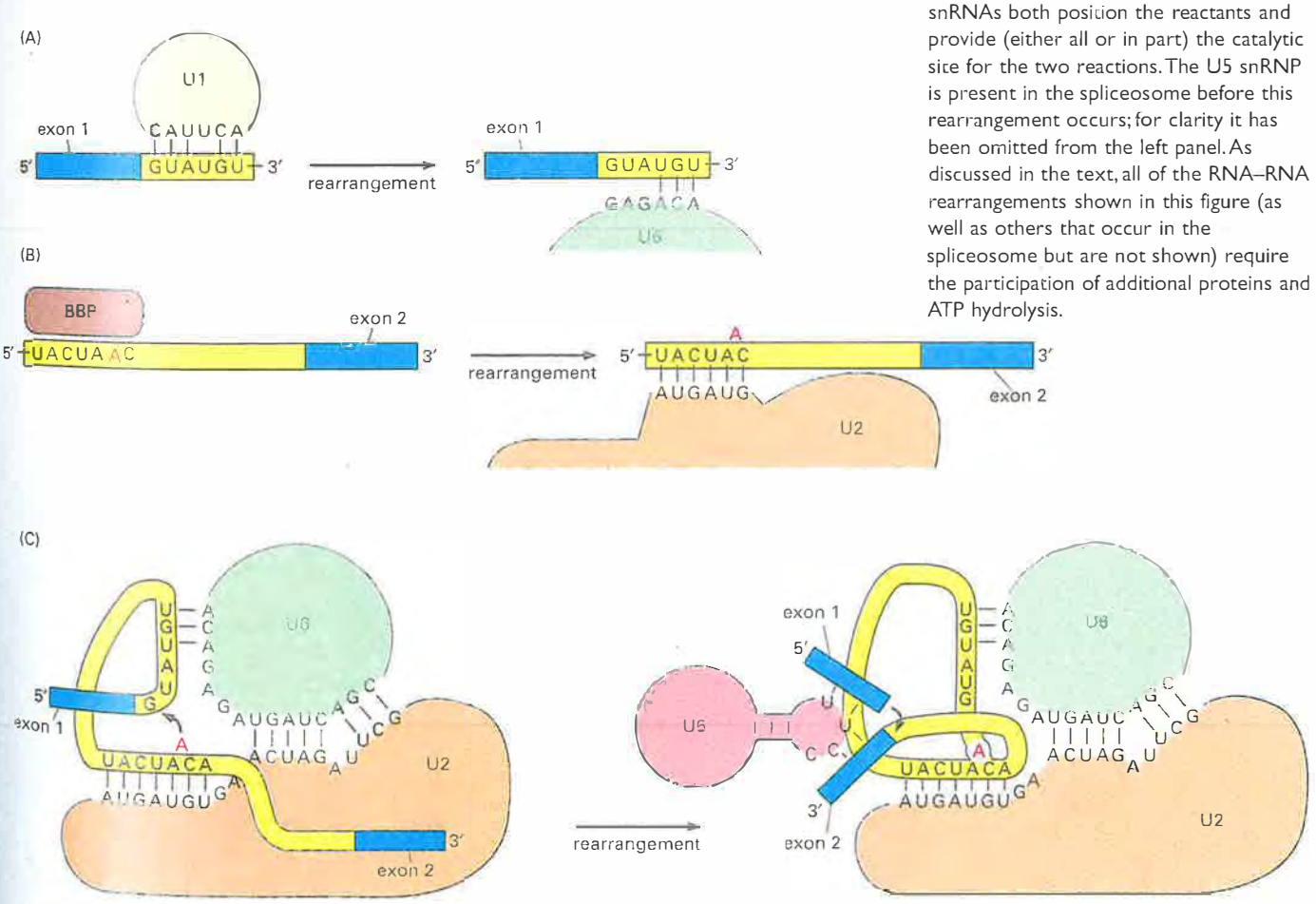


Figure 6–30 Several of the rearrangements that take place in the spliceosome during pre-mRNA splicing. Shown here are the details for the yeast *Saccharomyces cerevisiae*, in which the nucleotide sequences involved are slightly different from those in human cells. (A) The exchange of U1 snRNP for U6 snRNP occurs before the first phosphoryl-transfer reaction (see Figure 6–29). This exchange allows the 5′ splice site to be read by two different snRNPs, thereby increasing the accuracy of 5′ splice site selection by the spliceosome. (B) The branch-point site is first recognized by BBP and subsequently by U2 snRNP; as in (A), this "check and recheck" strategy provides increased accuracy of site selection. The binding of U2 to the branch-point forces the appropriate adenine (in *red*) to be unpaired and thereby activates it for the attack on the 5′ splice site (see Figure 6–29). This, in combination with recognition by BBP, is the way in which the spliceosome accurately chooses the adenine that is ultimately to form the branch point. (C) After the first phosphoryl-transfer reaction *(left)* has occurred, U5 snRNP undergoes a rearrangement that brings the two exons into close proximity for the second phosphoryl-transfer reaction *(right)*. The snRNAs both position the reactants and provide (either all or in part) the catalytic site for the two reactions. The U5 snRNP is present in the spliceosome before this rearrangement occurs; for clarity it has been omitted from the left panel. As discussed in the text, all of the RNA–RNA rearrangements shown in this figure (as well as others that occur in the spliceosome but are not shown) require the participation of additional proteins and ATP hydrolysis.

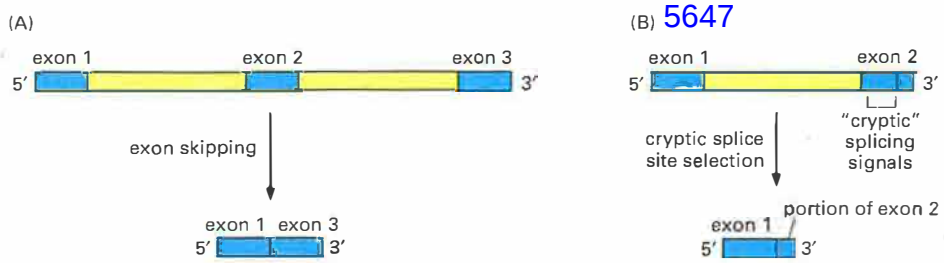


Figure 6–31 Two types of splicing errors. Both types might be expected to occur frequently if splice-site selection were performed by the spliceosome on a preformed, protein-free RNA molecule. "Cryptic" splicing signals are nucleotide sequences of RNA that closely resemble true splicing signals.

Once the splicing chemistry is completed, the snRNPs remain bound to the lariat and the spliced product is released. The disassembly of these snRNPs from the lariat (and from each other) requires another series of RNA–RNA rearrangements that require ATP hydrolysis, thereby returning the snRNAs to their original configuration so that they can be used again in a new reaction.

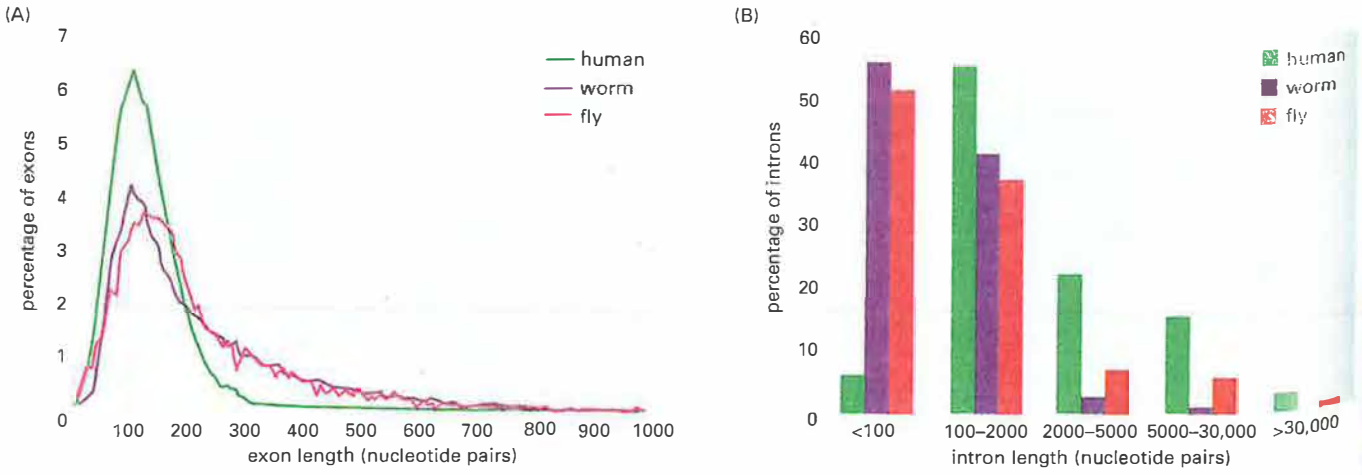
Ordering Influences in the Pre-mRNA Help to Explain How the Proper Splice Sites Are Chosen

As we have seen, intron sequences vary enormously in size, with some being in excess of 100,000 nucleotides. If splice-site selection were determined solely by the snRNPs acting on a preformed, protein-free RNA molecule, we would expect splicing mistakes—such as exon skipping and the use of cryptic splice sites—to be very common (Figure 6–31).

The fidelity mechanisms built into the spliceosome are supplemented by two additional factors that help ensure that splicing occurs accurately. These ordering influences in the pre-mRNA increase the probability that the appropriate pairs of 5′ and 3′ splice sites will be brought together in the spliceosome before the splicing chemistry begins. The first results from the assembly of the spliceosome occurring as the pre-mRNA emerges from a transcribing RNA polymerase II (see Figure 6–23). As for 5′ cap formation, several components of the spliceosome seem to be carried on the phosphorylated tail of RNA polymerase. Their transfer directly from the polymerase to the nascent pre-mRNA presumably helps the cell to keep track of introns and exons: the snRNPs at a 5′ splice site are initially presented with only a single 3′ splice site since the sites further downstream have not yet been synthesized. This feature helps to prevent inappropriate exon skipping.

The second factor that helps the cell to choose splice sites has been termed the "exon definition hypothesis," and it is understood only in outline. Exon size tends to be much more uniform than intron size, averaging about 150 nucleotide pairs across a wide variety of eucaryotic organisms (Figure 6–32). As RNA synthesis proceeds, a group of spliceosome components, called the SR proteins (so-named because they contain a domain rich in serines and arginines), are thought to assemble on exon sequences and mark off each 3′ and 5′ splice site starting at the 5′ end of the RNA (Figure 6–33). This assembly takes place in conjunction with the U1 snRNA, which marks one exon boundary, and U2AF,

Figure 6–32 Variation in intron and exon lengths in the human, worm, and fly genomes. (A) Size distribution of exons. (B) Size distribution of introns. Note that exon length is much more uniform than intron length. (Adapted from International Human Genome Sequencing Consortium, *Nature* 409:860–921, 2001.)



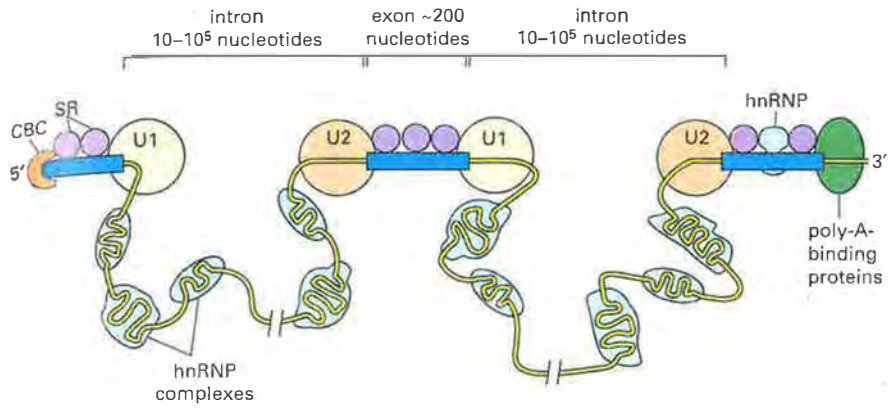


Figure 6–33 The exon definition hypothesis. According to one proposal, SR proteins bind to each exon sequence in the pre-mRNA and thereby help to guide the snRNPs to the proper intron/exon boundaries. This demarcation of exons by the SR proteins occurs co-transcriptionally, beginning at the CBC (cap-binding complex) at the 5′ end. As indicated, the intron sequences in the pre-mRNA, which can be extremely long, are packaged into hnRNP (heterogeneous nuclear ribonucleoprotein) complexes that compact them into more manageable structures and perhaps mask cryptic splice sites. Each hnRNP complex forms a particle approximately twice the diameter of a nucleosome, and the core is composed of a set of at least eight different proteins. It has been proposed that hnRNP proteins preferentially associate with intron sequences and that this preference also helps the spliceosome distinguish introns from exons. However, as shown, at least some hnRNP proteins may bind to exon sequences but their role, if any, in exon definition has yet to be established. (Adapted from R. Reed, *Curr. Opin. Cell Biol.* 12:340–345, 2000.)

which initially helps to specify the other. By specifically marking the exons in this way, the cell increases the accuracy with which the initial splicing components are deposited on the nascent RNA and thereby helps to avoid cryptic splice sites. How the SR proteins discriminate exon sequences from intron sequences is not understood; however, it is known that some of the SR proteins bind preferentially to RNA sequences in specific exons. In principle, the redundancy in the genetic code could have been exploited during evolution to select for binding sites for SR proteins in exons, allowing these sites to be created without constraining amino acid sequences.

Both the marking out of exon and intron boundaries and the assembly of the spliceosome begin on an RNA molecule while it is still being elongated by RNA polymerase at its 3′ end. However, the actual chemistry of splicing can take place much later. This delay means that intron sequences are not necessarily removed from a pre-mRNA molecule in the order in which they occur along the RNA chain. It also means that, although spliceosome assembly is co-transcriptional, the splicing reactions sometimes occur posttranscriptionally—that is, after a complete pre-mRNA molecule has been made.

A Second Set of snRNPs Splice a Small Fraction of Intron Sequences in Animals and Plants

Simple eucaryotes such as yeast have only one set of snRNPs that perform all pre-mRNA splicing. However, more complex eucaryotes such as flies, mammals, and plants have a second set of snRNPs that direct the splicing of a small fraction of their intron sequences. This minor form of spliceosome recognizes a different set of DNA sequences at the 5′ and 3′ splice junctions and at the branch point; it is called the *AT–AC spliceosome* because of the nucleotide sequence determinants at its intron–exon borders (Figure 6–34). Despite recognizing different nucleotide sequences, the snRNPs in this spliceosome make the same types of RNA–RNA interactions with the pre-mRNA and with each other as do the major snRNPs (Figure 6–34B). The recent discovery of this class of snRNPs gives us confidence in the base-pair interactions deduced for the major spliceosome, because it provides an independent set of molecules that undergo the same RNA–RNA interactions despite differences in the RNA sequences involved.

A particular variation on splicing, called **trans-splicing**, has been discovered in a few eucaryotic organisms. These include the single-celled trypanosomes—protozoans that cause African sleeping sickness in humans—and the model multicellular organism, the nematode worm. In trans-splicing, exons from two separate RNA transcripts are spliced together to form a mature mRNA molecule (see Figure 6–34). Trypanosomes produce all of their mRNAs in this way, whereas only about 1% of nematode mRNAs are produced by trans-splicing. In both cases, a single exon is spliced onto the 5′ end of many different RNA transcripts produced by the cell; in this way, all of the products of trans-splicing have the same 5′ exon and different 3′ exons. Many of the same snRNPs that function in conventional splicing are used in this reaction, although trans-splicing uses a unique snRNP (called the SL RNP) that brings in the common exon (see Figure 6–34).

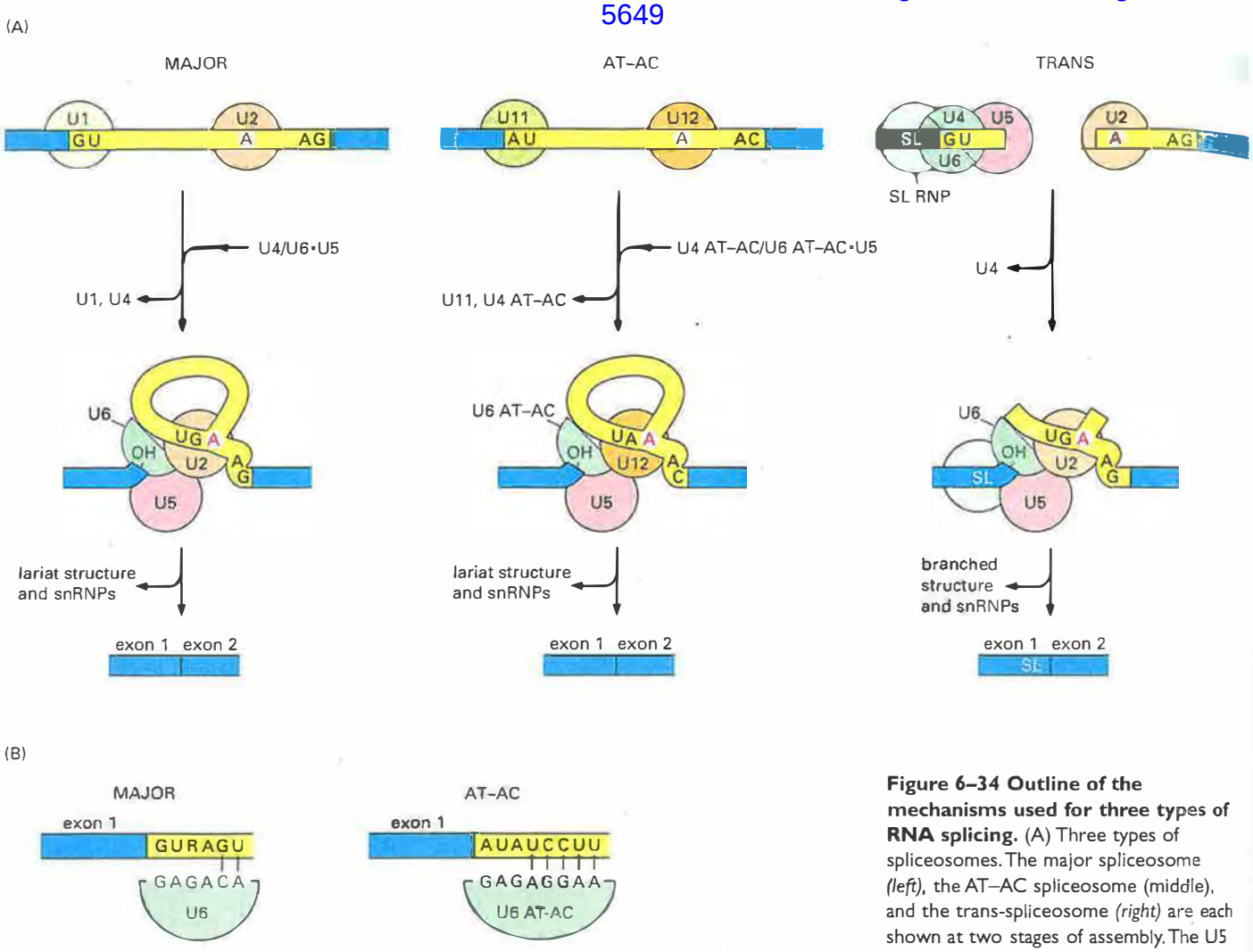


Figure 6–34 Outline of the mechanisms used for three types of RNA splicing. (A) Three types of spliceosomes. The major spliceosome *(left)*, the AT–AC spliceosome (middle), and the trans-spliceosome *(right)* are each shown at two stages of assembly. The U5 snRNP is the only component that is common to all three spliceosomes. Introns removed by the AT–AC spliceosome have a different set of consensus nucleotide sequences from those removed by the major spliceosome. In humans, it is estimated that 0.1% of introns are removed by the AT–AC spliceosome. In trans-splicing, the SL snRNP is consumed in the reaction because a portion of the SL snRNA becomes the first exon of the mature mRNA. (B) The major U6 snRNP and the U6 AT–AC snRNP both recognize the 5′ splice junction, but they do so through a different set of base-pair interactions. The sequences shown are from humans. (Adapted from Y.-T. Yu et al., The RNA World, pp. 487–524. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press, 1999.)

The reason that a few organisms use trans-splicing is not known; however, it is thought that the common 5′ exon may aid in the translation of the mRNA. Thus, the products of trans-splicing in nematodes seem to be translated with especially high efficiency.

RNA Splicing Shows Remarkable Plasticity

We have seen that the choice of splice sites depends on many features of the pre-mRNA transcript; these include the affinity of the three signals on the RNA (the 5′ and 3′ splice junctions and branch point) for the splicing machinery, the length and nucleotide sequence of the exon, the co-transcriptional assembly of the spliceosome, and the accuracy of the "bookkeeping" that underlies exon definition. So far we have emphasized the accuracy of the RNA splicing processes that occur in a cell. But it also seems that the mechanism has been selected for its flexibility, which allows the cell to try out new proteins on occasion. Thus, for example, when a mutation occurs in a nucleotide sequence critical for splicing of a particular intron, it does not necessarily prevent splicing of that intron altogether. Instead, the mutation typically creates a new pattern of splicing (Figure 6–35). Most commonly, an exon is simply skipped (Figure 6–35B). In other cases, the mutation causes a "cryptic" splice junction to be used (Figure 6–35C). Presumably, the splicing machinery has evolved to pick out the best possible pattern of splice junctions, and if the optimal one is damaged by mutation, it will seek out the next best pattern and so on. This flexibility in the process of RNA splicing suggests that changes in splicing patterns caused by random mutations have been an important pathway in the evolution of genes and organisms.

The plasticity of RNA splicing also means that the cell can easily regulate the pattern of RNA splicing. Earlier in this section we saw that alternative splicing

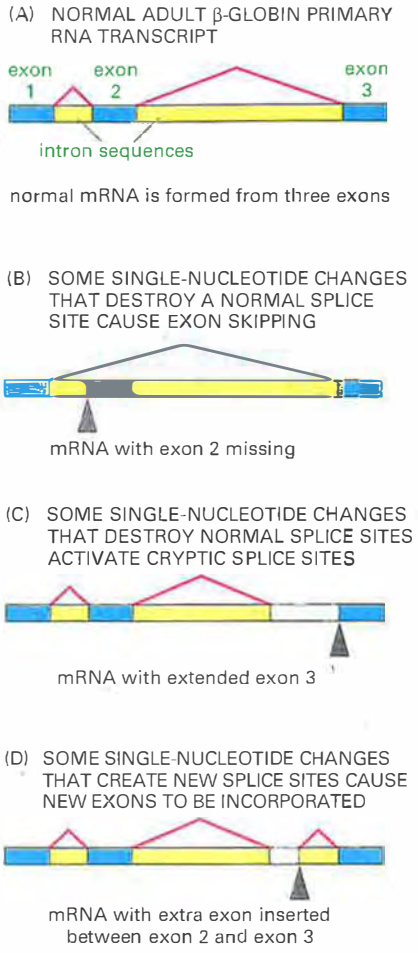


Figure 6–35 Abnormal processing of the β-globin primary RNA transcript in humans with the disease β thalassemia. In the examples shown, the disease is caused by splice-site mutations, denoted by *black arrowheads*. The *dark blue boxes* represent the three normal exon sequences; the *red lines* are used to indicate the 5′ and 3′ splice sites that are used in splicing the RNA transcript. The *light blue boxes* depict new nucleotide sequences included in the final mRNA molecule as a result of the mutation. Note that when a mutation leaves a normal splice site without a partner, an exon is skipped or one or more abnormal "cryptic" splice sites nearby is used as the partner site, as in (C) and (D). (Adapted in part from S.H. Orkin, in The Molecular Basis of Blood Diseases [G. Stamatoyannopoulos et al., eds.], pp. 106–126. Philadelphia: Saunders, 1987.)

can give rise to different proteins from the same gene. Some examples of alternative splicing are constitutive; that is, the alternatively spliced mRNAs are produced continuously by cells of an organism. However, in most cases, the splicing patterns are regulated by the cell so that different forms of the protein are produced at different times and in different tissues (see Figure 6–27). In Chapter 7 we return to this issue to discuss some specific examples of regulated RNA splicing.

Spliceosome-catalyzed RNA Splicing Probably Evolved from Self-splicing Mechanisms

When the spliceosome was first discovered, it puzzled molecular biologists. Why do RNA molecules instead of proteins perform important roles in splice site recognition and in the chemistry of splicing? Why is a lariat intermediate used rather than the apparently simpler alternative of bringing the 5′ and 3′ splice sites together in a single step, followed by their direct cleavage and rejoining? The answers to these questions reflect the way in which the spliceosome is believed to have evolved.

As discussed briefly in Chapter 1 (and taken up again in more detail in the final section of this chapter), it is thought that early cells used RNA molecules rather than proteins as their major catalysts and that they stored their genetic information in RNA rather than in DNA sequences. RNA-catalyzed splicing reactions presumably had important roles in these early cells. As evidence, some *self-splicing RNA* introns (that is, intron sequences in RNA whose splicing out can occur in the absence of proteins or any other RNA molecules) remain today—for example, in the nuclear rRNA genes of the ciliate *Tetrahymena*, in a few bacteriophage T4 genes, and in some mitochondrial and chloroplast genes.

A self-splicing intron sequence can be identified in a test tube by incubating a pure RNA molecule that contains the intron sequence and observing the splicing reaction. Two major classes of self-splicing intron sequences can be distinguished in this way. *Group I intron sequences* begin the splicing reaction by binding a G nucleotide to the intron sequence; this G is thereby activated to form the attacking group that will break the first of the phosphodiester bonds cleaved during splicing (the bond at the 5′ splice site). In *group II intron sequences*, an especially reactive A residue in the intron sequence is the attacking group, and a lariat intermediate is generated. Otherwise the reaction pathways for the two types of self-splicing intron sequences are the same. Both are presumed to represent vestiges of very ancient mechanisms (Figure 6–36).

For both types of self-splicing reactions, the nucleotide sequence of the intron is critical; the intron RNA folds into a specific three-dimensional structure, which brings the 5′ and 3′ splice junctions together and provides precisely positioned reactive groups to perform the chemistry (see Figure 6–6C). Based on the fact that the chemistries of their splicing reactions are so similar, it has been proposed that the pre-mRNA splicing mechanism of the spliceosome evolved from group II splicing. According to this idea, when the spliceosomal snRNPs took over the structural and chemical roles of the group II introns, the strict sequence constraints on intron sequences would have disappeared, thereby permitting a vast expansion in the number of different RNAs that could be spliced.

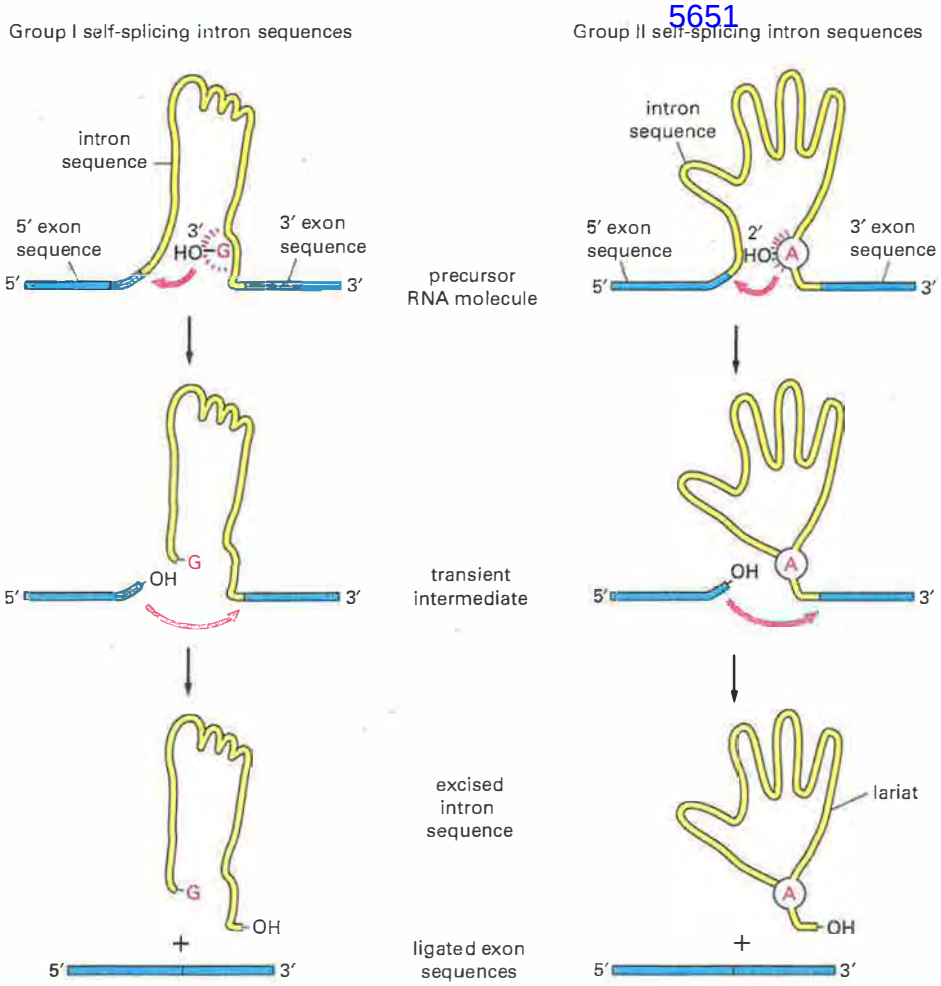


Figure 6–36 The two known classes of self-splicing intron sequences. The group I intron sequences bind a free G nucleotide to a specific site on the RNA to initiate splicing, while the group II intron sequences use an especially reactive A nucleotide in the intron sequence itself for the same purpose. The two mechanisms have been drawn to emphasize their similarities. Both are normally aided in the cell by proteins that speed up the reaction, but the catalysis is nevertheless mediated by the RNA in the intron sequence. Both types of self-splicing reactions require the intron to be folded into a highly specific three-dimensional structure that provides the catalytic activity for the reaction (see Figure 6–6). The mechanism used by group II intron sequences releases the intron as a lariat structure and closely resembles the pathway of pre-mRNA splicing catalyzed by the spliceosome (compare with Figure 6–29). The great majority of RNA splicing in eucaryotic cells is performed by the spliceosome, and self-splicing RNAs represent unusual cases. (Adapted from T.R. Cech, *Cell* 44:207–210, 1986.)

RNA-Processing Enzymes Generate the 3′ End of Eucaryotic mRNAs

As previously explained, the 5′ end of the pre-mRNA produced by RNA polymerase II is capped almost as soon as it emerges from the RNA polymerase. Then, as the polymerase continues its movement along a gene, the spliceosome components assemble on the RNA and delineate the intron and exon boundaries. The long C-terminal tail of the RNA polymerase coordinates these processes by transferring capping and splicing components directly to the RNA as the RNA emerges from the enzyme. As we see in this section, as RNA polymerase II terminates transcription at the end of a gene, it uses a similar mechanism to ensure that the 3′ end of the pre-mRNA becomes appropriately processed.

As might be expected, the 3′ ends of mRNAs are ultimately specified by DNA signals encoded in the genome (Figure 6–37). These DNA signals are transcribed

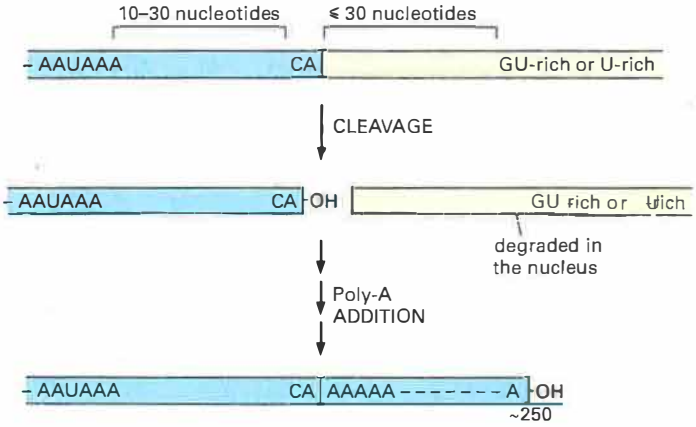


Figure 6–37 Consensus nucleotide sequences that direct cleavage and polyadenylation to form the 3′ end of a eucaryotic mRNA. These sequences are encoded in the genome and are recognized by specific proteins after they are transcribed into RNA. The hexamer AAUAAA is bound by CPSF, the GU-rich element beyond the cleavage site by CstF (see Figure 6–38), and the CA sequence by a third factor required for the cleavage step. Like other consensus nucleotide sequences discussed in this chapter (see Figure 6–12), the sequences shown in the figure represent a variety of individual cleavage and polyadenylation signals.

into RNA as the RNA polymerase II moves through them, and they are then recognized (as RNA) by a series of RNA-binding proteins and RNA-processing enzymes (Figure 6–38). Two multisubunit proteins, called CstF (cleavage stimulation factor F) and CPSF (cleavage and polyadenylation specificity factor), are of special importance. Both of these proteins travel with the RNA polymerase tail and are transferred to the 3′ end processing sequence on an RNA molecule as it emerges from the RNA polymerase. Some of the subunits of CPSF are associated with the general transcription factor TFIID, which, as we saw earlier in this chapter, is involved in transcription initiation. During transcription initiation, these subunits may be transferred from TFIID to the RNA polymerase tail, remaining associated there until the polymerase has transcribed through the end of a gene.

Once CstF and CPSF bind to specific nucleotide sequences on an emerging RNA molecule, additional proteins assemble with them to perform the processing that creates the 3′ end of the mRNA. First, the RNA is cleaved (see Figure 6–38). Next an enzyme called poly-A polymerase adds, one at a time, approximately 200 A nucleotides to the 3′ end produced by the cleavage. The nucleotide precursor for these additions is ATP, and the same type of 5′-to-3′ bonds are formed as in conventional RNA synthesis (see Figure 6–4). Unlike the usual RNA polymerases, poly-A polymerase does not require a template; hence the poly-A tail of eucaryotic mRNAs is not directly encoded in the genome. As the poly-A tail is synthesized, proteins called poly-A-binding proteins assemble onto it and, by a poorly understood mechanism, determine the final length of the tail. Poly-A-binding proteins remain bound to the poly-A tail as the mRNA makes its journey from the nucleus to the cytosol and they help to direct the synthesis of a protein on the ribosome, as we see later in this chapter.

After the 3′ end of a eucaryotic pre-mRNA molecule has been cleaved, the RNA polymerase II continues to transcribe, in some cases continuing as many as several hundred nucleotides beyond the DNA that contains the 3′ cleavage-site information. But the polymerase soon releases its grip on the template and transcription terminates; the piece of RNA downstream of the cleavage site is then degraded in the cell nucleus. It is not yet understood what triggers the loss in polymerase II processivity after the RNA is cleaved. One idea is that the transfer of the 3′ end processing factors from the RNA polymerase to the RNA causes a conformational change in the polymerase that loosens its hold on DNA; another is that the lack of a cap structure (and the CBC) on the 5′ end of the RNA that emerges from the polymerase somehow signals to the polymerase to terminate transcription.

Mature Eucaryotic mRNAs Are Selectively Exported from the Nucleus

We have seen how eucaryotic pre-mRNA synthesis and processing takes place in an orderly fashion within the cell nucleus. However, these events create a special problem for eucaryotic cells, especially those of complex organisms where the introns are vastly longer than the exons. Of the pre-mRNA that is synthesized, only a small fraction—the mature mRNA—is of further use to the cell. The rest—excised introns, broken RNAs, and aberrantly spliced pre-mRNAs—is not only useless but could be dangerous if it was not destroyed. How then does the cell distinguish between the relatively rare mature mRNA molecules it wishes to keep and the overwhelming amount of debris from RNA processing? The answer is that transport of mRNA from the nucleus to the cytoplasm, where it is translated into protein, is highly selective—being closely coupled to correct RNA processing. This coupling is achieved by the *nuclear pore complex*, which recognizes and transports only completed mRNAs.

We have seen that as a pre-mRNA molecule is synthesized and processed, it is bound by a variety of proteins, including the cap-binding complex, the SR proteins, and the poly-A binding proteins. To be "export-ready," it seems than an mRNA must be bound by the appropriate set of proteins—with certain proteins such as the cap-binding complex being present, and others such as snRNP proteins absent. Additional proteins, placed on the RNA during splicing, seem to

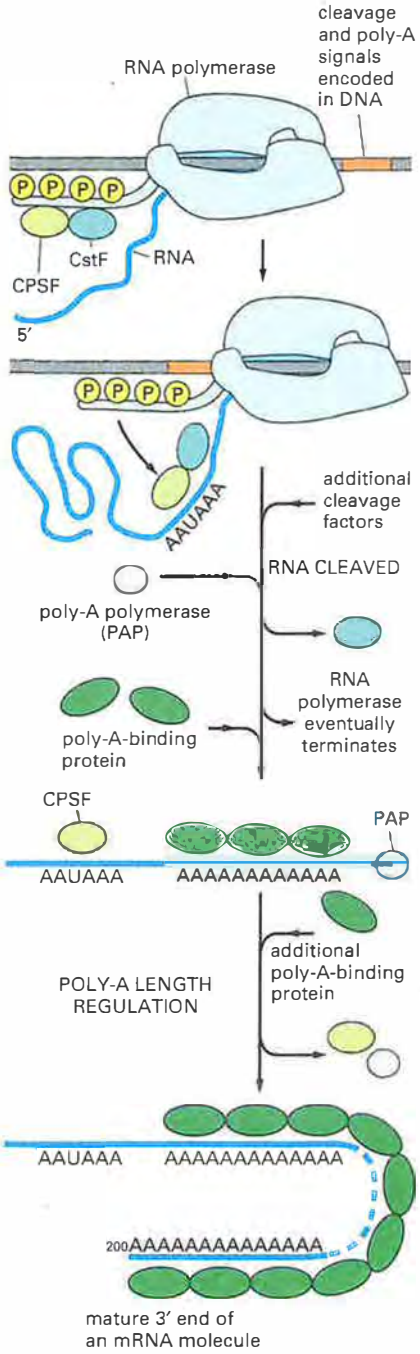


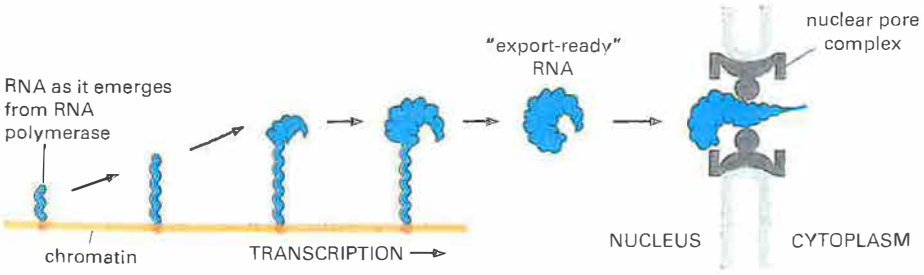
Figure 6–38 Some of the major steps in generating the 3′ end of a eucaryotic mRNA. This process is much more complicated than the analogous process in bacteria, where the RNA polymerase simply stops at a termination signal and releases both the 3′ end of its transcript and the DNA template (see Figure 6–10).

mark exon-exon boundaries and thereby signify completed splicing events. Only if the proper set of proteins is bound to an mRNA is it guided through the **nuclear pore complex** into the cytosol. As described in Chapter 12, nuclear pore complexes are aqueous channels in the nuclear membrane that directly connect the nucleoplasm and cytosol. Small molecules (less than 50,000 daltons) can diffuse freely through them. However, most of the macromolecules in cells, including mRNAs complexed with proteins, are far too large to pass through the pores without a special process to move them. An active transport of substances through the nuclear pore complexes occurs in both directions. As explained in Chapter 12, signals on the macromolecule determine whether it is exported from the nucleus (a mRNA, for example) or imported into it (an RNA polymerase, for example). For the case of mRNAs, the bound proteins that mark completed splicing events are of particular importance, as they are known to serve directly as RNA export factors (see Figure 12–16). mRNAs transcribed from genes that lack introns apparently contain nucleotide sequences that are directly recognized by other RNA export factors. Eucaryotic cells thus use their nuclear pore complexes as gates that allow only useful RNA molecules to enter the cytoplasm.

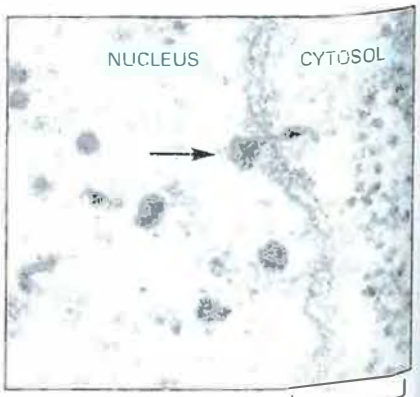
Of all the proteins that assemble on pre-mRNA molecules as they emerge from transcribing RNA polymerases, the most abundant are the hnRNPs (heterogeneous nuclear ribonuclear proteins). Some of these proteins (there are approximately 30 of them in humans) remove the hairpin helices from the RNA so that splicing and other signals on the RNA can be read more easily. Others package the RNA contained in the very long intron sequences typically found in genes of complex organisms (see Figure 6–33). Apart from histones, certain hnRNP proteins are the most abundant proteins in the cell nucleus, and they may play a particularly important role in distinguishing mature mRNA from processing debris. hnRNP particles (nucleosome-like complexes of hnRNP proteins and RNA—see Figure 6–33) are largely excluded from exon sequences, perhaps by prior binding of spliceosome components. They remain on excised introns and probably help mark them for nuclear retention and eventual destruction.

The export of mRNA–protein complexes from the nucleus can be observed with an electron microscope for the unusually abundant mRNA of the insect Balbiani Ring genes. As these genes are transcribed, the newly formed RNA is seen to be packaged by proteins (including hnRNP and SR proteins). This protein–RNA complex undergoes a series of structural transitions, probably reflecting RNA processing events, culminating in a curved fiber (Figure 6–39). This curved fiber then moves through the nucleoplasm and enters the nuclear pore complex (with its 5′ cap proceeding first), and it undergoes another series of structural transitions as it moves through the NPC. These and other observations reveal that the pre-mRNA–protein and mRNA–protein complexes are dynamic structures that gain and lose numerous specific proteins during RNA synthesis, processing, export, and translation (Figure 6–40).

Before discussing what happens to mRNAs after they leave the nucleus, we briefly consider how the synthesis and processing of noncoding RNA molecules



(A)



(B)

Figure 6–39 Transport of a large mRNA molecule through the nuclear pore complex. (A) The maturation of a Balbiani Ring mRNA molecule as it is synthesized by RNA polymerase and packaged by a variety of nuclear proteins. This drawing of unusually abundant RNA produced by an insect cell is based on EM micrographs such as that shown in (B). Balbiani Rings are described in Chapter 4. (A, adapted from B. Daneholt, *Cell* 88:585–588, 1997; B, from B.J. Stevens and H. Swift, *J. Cell Biol.* 31:55–77, 1966. © The Rockefeller University Press.)

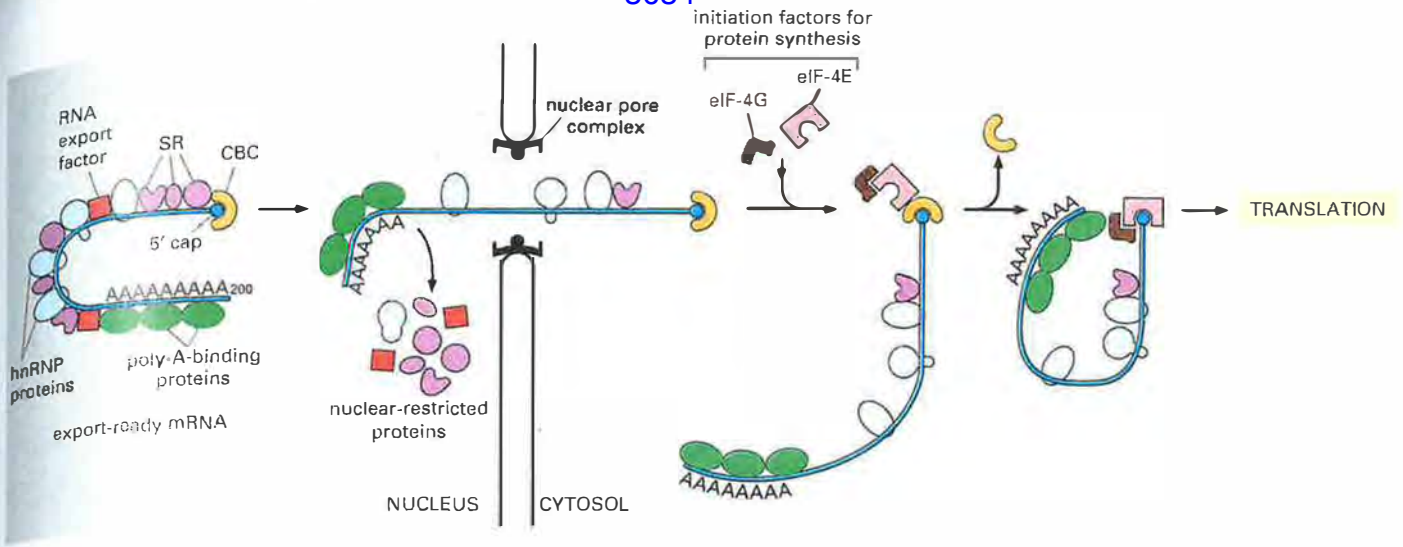


Figure 6–40 Schematic illustration of an "export-ready" mRNA molecule and its transport through the nuclear pore. As indicated, some proteins travel with the mRNA as it moves through the pore, whereas others remain in the nucleus. Once in the cytoplasm, the mRNA continues to shed previously bound proteins and acquire new ones; these substitutions affect the subsequent translation of the message. Because some are transported with the RNA, the proteins that become bound to an mRNA in the nucleus can influence its subsequent stability and translation in the cytosol. RNA export factors, shown in the nucleus, play an active role in transporting the mRNA to the cytosol (see Figure 12–16). Some are deposited at exon-exon boundaries as splicing is completed, thus signifying those regions of the RNA that have been properly spliced.

occurs. Although there are many other examples, our discussion focuses on the rRNAs that are critically important for the translation of mRNAs into protein.

Many Noncoding RNAs Are Also Synthesized and Processed in the Nucleus

A few per cent of the dry weight of a mammalian cell is RNA; of that, only about 3–5% is mRNA. A fraction of the remainder represents intron sequences before they have been degraded, but most of the RNA in cells performs structural and catalytic functions (see Table 6–1, p. 306). The most abundant RNAs in cells are the ribosomal RNAs (rRNAs)—constituting approximately 80% of the RNA in rapidly dividing cells. As discussed later in this chapter, these RNAs form the core of the ribosome. Unlike bacteria—in which all RNAs in the cell are synthesized by a single RNA polymerase—eucaryotes have a separate, specialized polymerase, RNA polymerase I, that is dedicated to producing rRNAs. RNA polymerase I is similar structurally to the RNA polymerase II discussed previously; however, the absence of a C-terminal tail in polymerase I helps to explain why its transcripts are neither capped nor polyadenylated. As discussed earlier, this difference helps the cell distinguish between noncoding RNAs and mRNAs.

Because multiple rounds of translation of each mRNA molecule can provide an enormous amplification in the production of protein molecules, many of the proteins that are very abundant in a cell can be synthesized from genes that are present in a single copy per haploid genome. In contrast, the RNA components of the ribosome are final gene products, and a growing mammalian cell must synthesize approximately 10 million copies of each type of ribosomal RNA in each cell generation to construct its 10 million ribosomes. Adequate quantities of ribosomal RNAs can be produced only because the cell contains multiple copies of the **rRNA genes** that code for ribosomal RNAs (**rRNAs**). Even *E. coli* needs seven copies of its rRNA genes to meet the cell's need for ribosomes. Human cells contain about 200 rRNA gene copies per haploid genome, spread out in small clusters on five different chromosomes (see Figure 4–11), while cells

Figure 6–41 Transcription from tandemly arranged rRNA genes, as seen in the electron microscope. The pattern of alternating transcribed gene and nontranscribed spacer is readily seen. A higher-magnification view was shown in Figure 6–9. (From V.E. Foe, *Cold Spring Harbor Symp. Quant. Biol.* 42:723–740, 1978.)

of the frog *Xenopus* contain about 600 rRNA gene copies per haploid genome in a single cluster on one chromosome (Figure 6–41).

There are four types of eucaryotic rRNAs, each present in one copy per ribosome. Three of the four rRNAs (18S, 5.8S, and 28S) are made by chemically modifying and cleaving a single large precursor rRNA (Figure 6–42); the fourth (5S RNA) is synthesized from a separate cluster of genes by a different polymerase, RNA polymerase III, and does not require chemical modification. It is not known why this one RNA is transcribed separately.

Extensive chemical modifications occur in the 13,000-nucleotide-long precursor rRNA before the rRNAs are cleaved out of it and assembled into ribosomes. These include about 100 methylations of the 2′-OH positions on nucleotide sugars and 100 isomerizations of uridine nucleotides to pseudouridine (Figure 6–43A). The functions of these modifications are not understood in detail, but they probably aid in the folding and assembly of the final rRNAs and may also subtly alter the function of ribosomes. Each modification is made at a specific position in the precursor rRNA. These positions are specified by several hundred "guide RNAs," which locate themselves through base-pairing to the precursor rRNA and thereby bring an RNA-modifying enzyme to the appropriate position (Figure 6–43B). Other guide RNAs promote cleavage of the precursor rRNAs into the mature rRNAs, probably by causing conformational changes in the precursor rRNA. All of these guide RNAs are members of a large class of RNAs called **small nucleolar RNAs** (or **snoRNAs**), so named because these RNAs perform their functions in a subcompartment of the nucleus called the nucleolus. Many snoRNAs are encoded in the introns of other genes,

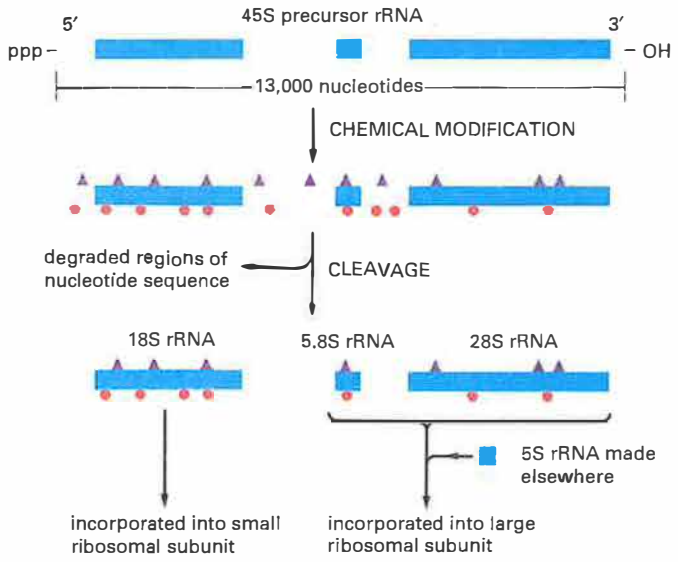


Figure 6–42 The chemical modification and nucleolytic processing of a eucaryotic 45S precursor rRNA molecule into three separate ribosomal RNAs. As indicated, two types of chemical modifications (shown in Figure 6–43) are made to the precursor rRNA before it is cleaved. Nearly half of the nucleotide sequences in this precursor rRNA are discarded and degraded in the nucleus. The rRNAs are named according to their "S" values, which refer to their rate of sedimentation in an ultra-centrifuge. The larger the S value, the larger the rRNA.

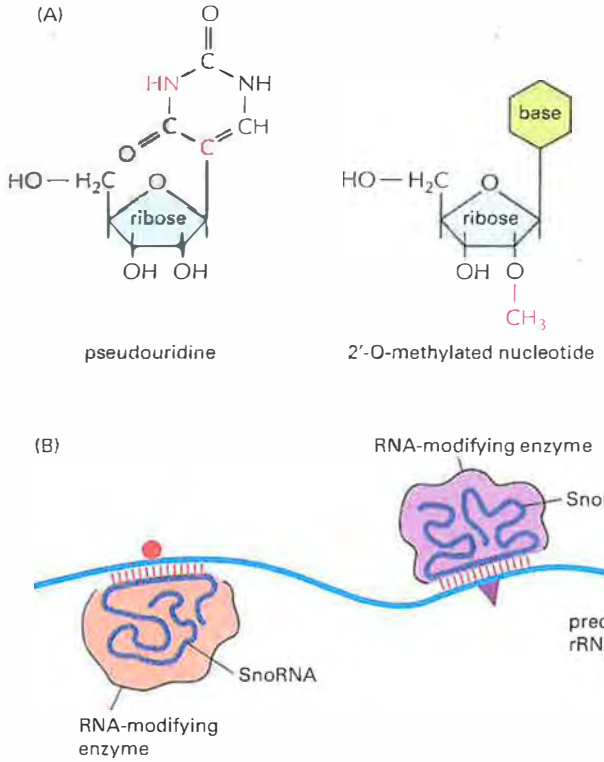


Figure 6–43 Modifications of the precursor rRNA by guide RNAs. (A) Two prominent covalent modifications occur after rRNA synthesis; the differences from the initially incorporated nucleotide are indicated by *red* atoms. (B) As indicated, snoRNAs locate the sites of modification by base-pairing to complementary sequences on the precursor rRNA. The snoRNAs are bound to proteins, and the complexes are called snoRNPs. snoRNPs contain the RNA modification activities, presumably contributed by the proteins but possibly by the snoRNAs themselves.

especially those encoding ribosomal proteins. They are therefore synthesized by RNA polymerase II and processed from excised intron sequences.

The Nucleolus Is a Ribosome-Producing Factory

The nucleolus is the most obvious structure seen in the nucleus of a eucaryotic cell when viewed in the light microscope. Consequently, it was so closely scrutinized by early cytologists that an 1898 review could list some 700 references. We now know that the nucleolus is the site for the processing of rRNAs and their assembly into ribosomes. Unlike other organelles in the cell, it is not bound by a membrane (Figure 6–44); instead, it is a large aggregate of macromolecules, including the rRNA genes themselves, precursor rRNAs, mature rRNAs, rRNA-processing enzymes, snoRNPs, ribosomal protein subunits and partly assembled

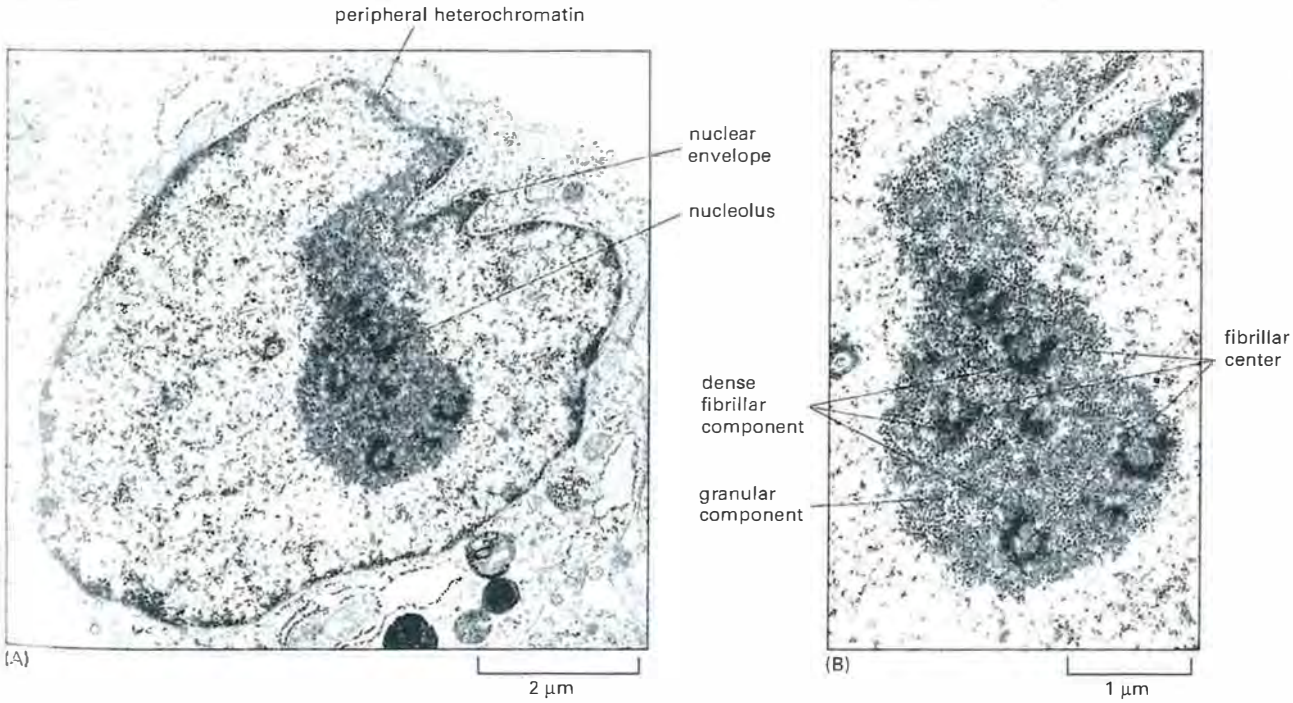
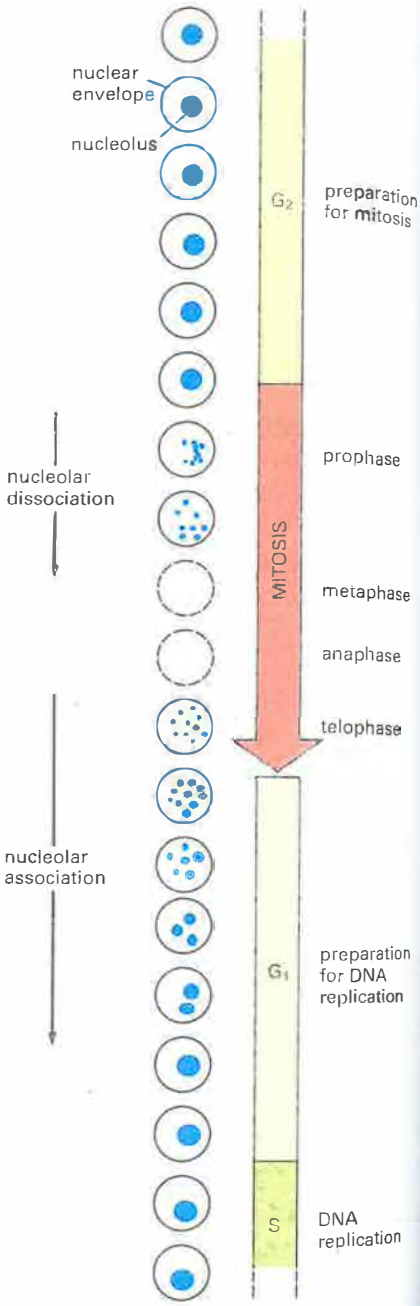


Figure 6–44 Electron micrograph of a thin section of a nucleolus in a human fibroblast, showing its three distinct zones. (A) View of entire nucleus. (B) High-power view of the nucleolus. It is believed that transcription of the rRNA genes takes place between the fibrillar center and the dense fibrillar component and that processing of the rRNAs and their assembly into ribosomes proceeds outward from the dense fibrillar component to the surrounding granular components. (Courtesy of E.G. Jordan and J. McGovern.)

Figure 6–45 Changes in the appearance of the nucleolus in a human cell during the cell cycle. Only the cell nucleus is represented in this diagram. In most eucaryotic cells the nuclear membrane breaks down during mitosis, as indicated by the dashed circles.



ribosomes. The close association of all these components presumably allows the assembly of ribosomes to occur rapidly and smoothly.

It is not yet understood how the nucleolus is held together and organized, but various types of RNA molecules play a central part in its chemistry and structure, suggesting that the nucleolus may have evolved from an ancient structure present in cells dominated by RNA catalysis. In present-day cells, the rRNA genes also have an important role in forming the nucleolus. In a diploid human cell, the rRNA genes are distributed into 10 clusters, each of which is located near the tip of one of the two copies of five different chromosomes (see Figure 4–11). Each time a human cell undergoes mitosis, the chromosomes disperse and the nucleolus breaks up; after mitosis, the tips of the 10 chromosomes coalesce as the nucleolus reforms (Figures 6–45 and 6–46). The transcription of the rRNA genes by RNA polymerase I is necessary for this process.

As might be expected, the size of the nucleolus reflects the number of ribosomes that the cell is producing. Its size therefore varies greatly in different cells and can change in a single cell, occupying 25% of the total nuclear volume in cells that are making unusually large amounts of protein.

A schematic diagram of the assembly of ribosomes is shown in Figure 6–47. In addition to its important role in ribosome biogenesis, the nucleolus is also the site where other RNAs are produced and other RNA–protein complexes are assembled. For example, the U6 snRNP, which, as we have seen, functions in pre-mRNA splicing (see Figure 6–29), is composed of one RNA molecule and at least seven proteins. The U6 snRNA is chemically modified by snoRNAs in the nucleolus before its final assembly there into the U6 snRNP. Other important RNA protein complexes, including telomerase (encountered in Chapter 5) and the signal recognition particle (which we discuss in Chapter 12), are also believed to be assembled at the nucleolus. Finally, the tRNAs (transfer RNAs) that carry the amino acids for protein synthesis are processed there as well. Thus, the nucleolus can be thought of as a large factory at which many different noncoding RNAs are processed and assembled with proteins to form a large variety of ribonucleoprotein complexes.

The Nucleus Contains a Variety of Subnuclear Structures

Although the nucleolus is the most prominent structure in the nucleus, several other nuclear bodies have been visualized and studied (Figure 6–48). These include Cajal bodies (named for the scientist who first described them in 1906), GEMS (Gemini of coiled bodies), and interchromatin granule clusters (also called "speckles"). Like the nucleolus, these other nuclear structures lack membranes

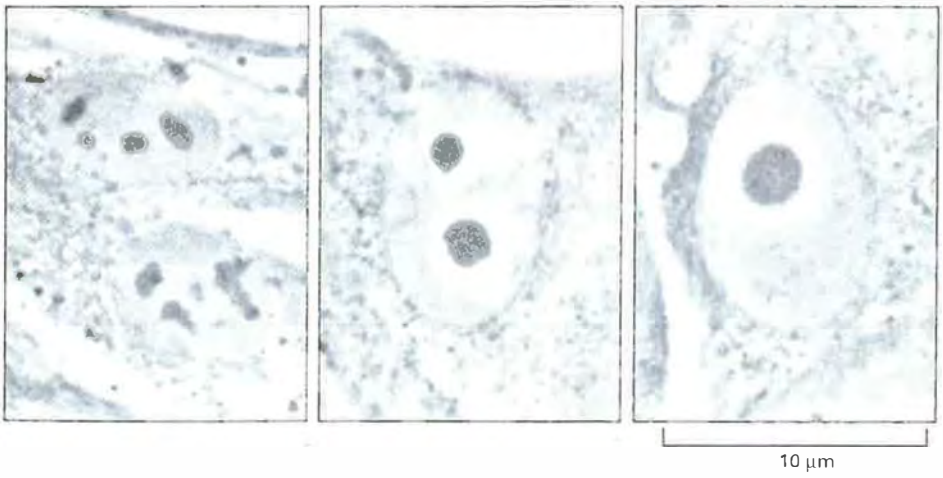


Figure 6–46 Nucleolar fusion. These light micrographs of human fibroblasts grown in culture show various stages of nucleolar fusion. After mitosis, each of the ten human chromosomes that carry a cluster of rRNA genes begins to form a tiny nucleolus, but these rapidly coalesce as they grow to form the single large nucleolus typical of many interphase cells. (Courtesy of E.G. Jordan and J. McGovern.)

and are highly dynamic; their appearance is probably the result of the tight association of protein and RNA (and perhaps DNA) components involved in the synthesis, assembly, and storage of macromolecules involved in gene expression. Cajal bodies and GEMS resemble one another and are frequently paired in the nucleus; it is not clear whether they are truly distinct structures. They may be sites where snRNAs and snoRNAs undergo their final modifications and assembly with protein. Both the RNAs and the proteins that make up the snRNPs are partly assembled in the cytoplasm, but they are transported into the nucleus for their final modifications. It has been proposed that Cajal bodies/GEMS are also sites where the snRNPs are recycled and their RNAs are "reset" after the rearrangements that occur during splicing (see p. 322). In contrast, the interchromatin granule clusters have been proposed to be stockpiles of fully mature snRNPs that are ready to be used in splicing of pre-mRNAs (Figure 6–49).

Scientists have had difficulties in working out the function of the small subnuclear structures just described. Much of the progress now being made depends on genetic tools—examination of the effects of designed mutations in mice or of spontaneous mutations in humans. As one example, GEMS contain the SMN (survival of motor neurons) protein. Certain mutations of the gene encoding this protein are the cause of inherited spinal muscular atrophy, a human disease characterized by a wasting away of the muscles. The disease seems

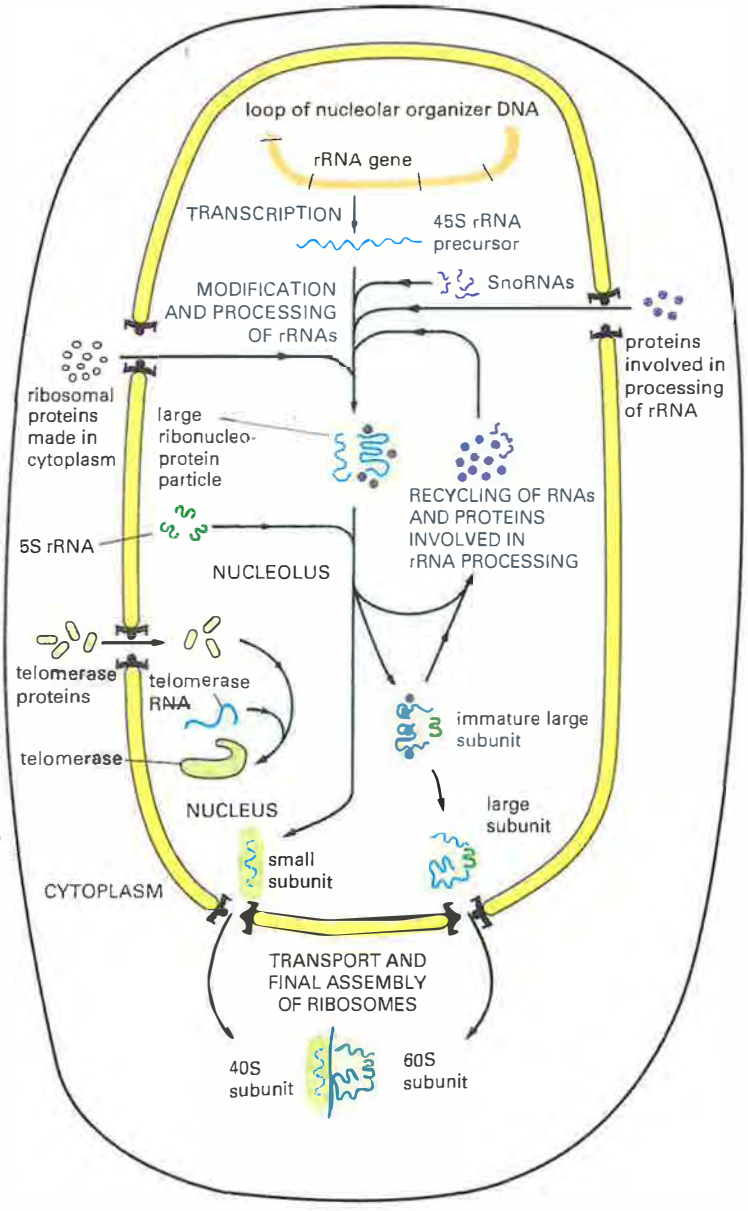


Figure 6–47 The function of the nucleolus in ribosome and other ribonucleoprotein synthesis. The 45S precursor rRNA is packaged in a large ribonucleoprotein particle containing many ribosomal proteins imported from the cytoplasm. While this particle remains in the nucleolus, selected pieces are added and others discarded as it is processed into immature large and small ribosomal subunits. The two ribosomal subunits are thought to attain their final functional form only as each is individually transported through the nuclear pores into the cytoplasm. Other ribonucleoprotein complexes, including telomerase shown here, are also assembled in the nucleolus.

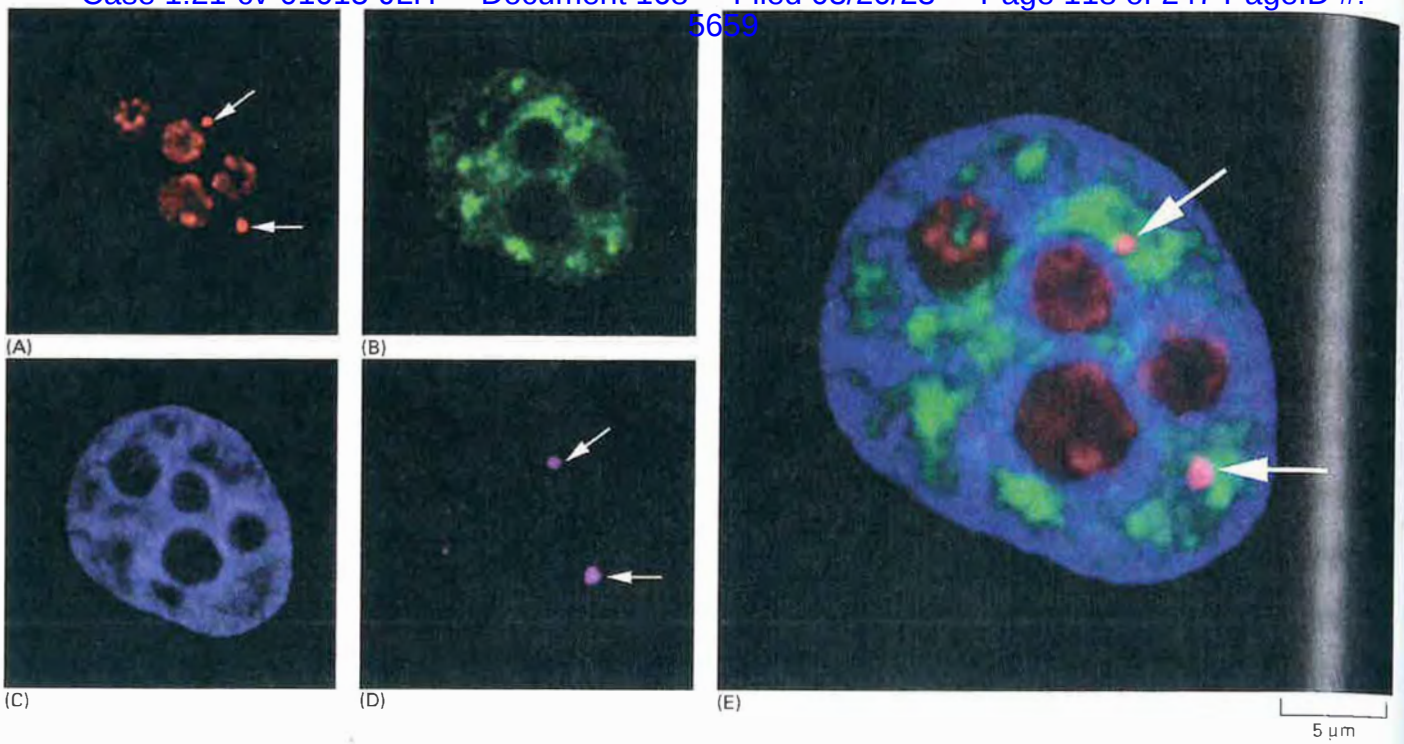


Figure 6–48 Visualization of chromatin and nuclear bodies. (A)–(D) show micrographs of the same human cell nucleus, each processed differently to show a particular set of nuclear structures. (E) shows an enlarged superposition of all four individual images. (A) shows the location of the protein fibrillarin (a component of several snoRNPs), which is present at both nucleoli and Cajal bodies, the latter indicated by arrows. (B) shows interchromatin granule clusters or "speckles" detected by using antibodies against a protein involved in pre-mRNA splicing. (C) is stained to show bulk chromatin. (D) shows the location of the protein coilin, which is present at Cajal bodies (indicated by arrows). (From J.R. Swedlow and A.I. Lamond, *Gen. Biol.* 2:1–7, 2001; micrographs courtesy of Judith Sleeman.)

to be caused by a subtle defect in snRNP assembly and subsequent pre-mRNA splicing. More severe defects would be expected to be lethal.

Given the importance of nuclear subdomains in RNA processing, it might have been expected that pre-mRNA splicing would occur in a particular location in the nucleus, as it requires numerous RNA and protein components. However,

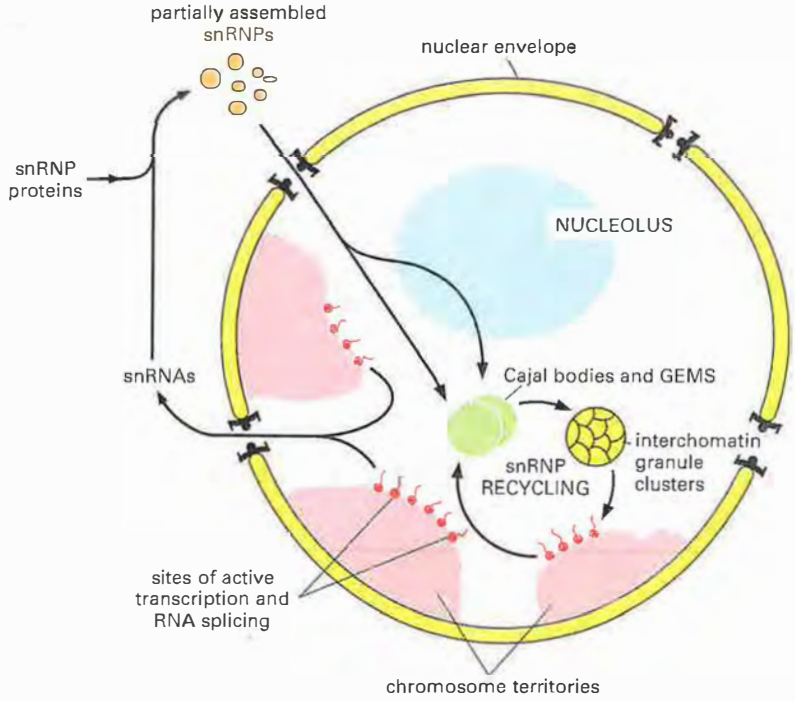


Figure 6–49 Schematic view of subnuclear structures. A typical vertebrate nucleus has several Cajal bodies, which are proposed to be the sites where snRNPs and snoRNPs undergo their final modifications. Interchromatin granule clusters are proposed to be storage sites for fully mature snRNPs. A typical vertebrate nucleus has 20–50 interchromatin granule clusters.

After their initial synthesis, snRNAs are exported from the nucleus, after which they undergo 5′ and 3′ end-processing and assemble with the seven common snRNP proteins (called Sm proteins). These complexes are reimported into the nucleus and the snRNPs undergo their final modification in Cajal bodies. In addition, the U6 snRNP requires chemical modification by snoRNAs in the nucleolus. The sites of active transcription and splicing (approximately 2000–3000 sites per vertebrate nucleus) correspond to the "perichromatin fibers" seen under the electron microscope. (Adapted from J.D. Lewis and D. Tollervey, *Science* 288:1385–1389, 2000.)

we have seen that the assembly of splicing components on pre-mRNA is co-transcriptional; thus splicing must occur at many locations along chromosomes. We saw in Chapter 4 that interphase chromosomes occupy discrete territories in the nucleus, and transcription and pre-mRNA splicing must take place within these territories. However, interphase chromosomes are themselves dynamic and their exact positioning in the nucleus correlates with gene expression. For example, transcriptionally silent regions of interphase chromosomes are often associated with the nuclear envelope where the concentration of heterochromatin components is believed to be especially high. When these same regions become transcriptionally active, they relocate towards the interior of the nucleus, which is richer in the components required for mRNA synthesis. It has been proposed that, although a typical mammalian cell may be expressing on the order of 15,000 genes, transcription and RNA splicing may be localized to only several thousand sites in the nucleus. These sites themselves are highly dynamic and probably result from the association of transcription and splicing components to create small "assembly lines" where the local concentration of these components is very high. As a result, the nucleus seems to be highly organized into subdomains, with snRNPs, snoRNPs, and other nuclear components moving between them in an orderly fashion according to the needs of the cell (Figure 6–49).

Summary

Before the synthesis of a particular protein can begin, the corresponding mRNA molecule must be produced by transcription. Bacteria contain a single type of RNA polymerase (the enzyme that carries out the transcription of DNA into RNA). An mRNA molecule is produced when this enzyme initiates transcription at a promoter, synthesizes the RNA by chain elongation, stops transcription at a terminator, and releases both the DNA template and the completed mRNA molecule. In eucaryotic cells, the process of transcription is much more complex, and there are three RNA polymerases—designated polymerase I, II, and III—that are related evolutionarily to one another and to the bacterial polymerase.

Eucaryotic mRNA is synthesized by RNA polymerase II. This enzyme requires a series of additional proteins, termed the general transcription factors, to initiate transcription on a purified DNA template and still more proteins (including chromatin-remodeling complexes and histone acetyltransferases) to initiate transcription on its chromatin template inside the cell. During the elongation phase of transcription, the nascent RNA undergoes three types of processing events: a special nucleotide is added to its 5′ end (capping), intron sequences are removed from the middle of the RNA molecule (splicing), and the 3′ end of the RNA is generated (cleavage and polyadenylation). Some of these RNA processing events that modify the initial RNA transcript (for example, those involved in RNA splicing) are carried out primarily by special small RNA molecules.

For some genes, RNA is the final product. In eucaryotes, these genes are usually transcribed by either RNA polymerase I or RNA polymerase III. RNA polymerase I makes the ribosomal RNAs. After their synthesis as a large precursor, the rRNAs are chemically modified, cleaved, and assembled into ribosomes in the nucleolus—a distinct subnuclear structure that also helps to process some smaller RNA–protein complexes in the cell. Additional subnuclear structures (including Cajal bodies and interchromatin granule clusters) are sites where components involved in RNA processing are assembled, stored, and recycled.

FROM RNA TO PROTEIN

In the preceding section we have seen that the final product of some genes is an RNA molecule itself, such as those present in the snRNPs and in ribosomes. However, most genes in a cell produce mRNA molecules that serve as intermediaries on the pathway to proteins. In this section we examine how the cell converts the information carried in an mRNA molecule into a protein molecule. This feat of translation first attracted the attention of biologists in the late 1950s,

GCA	AGA								UUA				AGC						
	AGG								UUG				AGU						
GCC	CGA						GGA		CUA				UCA	ACA			GUA		
GCG	CGC						GGC		CUC			CCA	UCC	ACC			GUC		UAA
GCU	CGG	GAC	AAC	UGC	GAA	CAA	GGG	CAC	AUA	CUG	AAA	CCC	UCG	ACG		UAC	GUG		UAG
	CGU	GAU	AAU	UGU	GAG	CAG	GGU	CAU	AUC	CUU	AAG	AUG	CCG	UCU	ACU	UGG	UAU	GUU	UGA
									AUU				UUC	CCU					
Ala	Arg	Asp	Asn	Cys	Glu	Gln	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val	stop
A	R	D	N	C	E	Q	G	H	I	L	K	M	F	P	S	T	W	Y	V	

when it was posed as the "coding problem": how is the information in a linear sequence of nucleotides in RNA translated into the linear sequence of a chemically quite different set of subunits—the amino acids in proteins? This fascinating question stimulated great excitement among scientists at the time. Here was a cryptogram set up by nature that, after more than 3 billion years of evolution, could finally be solved by one of the products of evolution—human beings. And indeed, not only has the code been cracked step by step, but in the year 2000 the elaborate machinery by which cells read this code—the ribosome—was finally revealed in atomic detail.

Figure 6–50 The genetic code. The standard one-letter abbreviation for each amino acid is presented below its three-letter abbreviation (see Panel 3–1, pp. 132–133, for the full name of each amino acid and its structure). By convention, codons are always written with the 5′-terminal nucleotide to the left. Note that most amino acids are represented by more than one codon, and that there are some regularities in the set of codons that specifies each amino acid. Codons for the same amino acid tend to contain the same nucleotides at the first and second positions, and vary at the third position. Three codons do not specify any amino acid but act as termination sites (stop codons), signaling the end of the protein-coding sequence. One codon—AUG—acts both as an initiation codon, signaling the start of a protein-coding message, and also as the codon that specifies methionine.

An mRNA Sequence Is Decoded in Sets of Three Nucleotides

Once an mRNA has been produced, by transcription and processing the information present in its nucleotide sequence is used to synthesize a protein. Transcription is simple to understand as a means of information transfer: since DNA and RNA are chemically and structurally similar, the DNA can act as a direct template for the synthesis of RNA by complementary base-pairing. As the term *transcription* signifies, it is as if a message written out by hand is being converted, say, into a typewritten text. The language itself and the form of the message do not change, and the symbols used are closely related.

In contrast, the conversion of the information in RNA into protein represents a **translation** of the information into another language that uses quite different symbols. Moreover, since there are only four different nucleotides in mRNA and twenty different types of amino acids in a protein, this translation cannot be accounted for by a direct one-to-one correspondence between a nucleotide in RNA and an amino acid in protein. The nucleotide sequence of a gene, through the medium of mRNA, is translated into the amino acid sequence of a protein by rules that are known as the **genetic code**. This code was deciphered in the early 1960s.

The sequence of nucleotides in the mRNA molecule is read consecutively in groups of three. RNA is a linear polymer of four different nucleotides, so there are $4 \times 4 \times 4 = 64$ possible combinations of three nucleotides: the triplets AAA, AUA, AUG, and so on. However, only 20 different amino acids are commonly found in proteins. Either some nucleotide triplets are never used, or the code is redundant and some amino acids are specified by more than one triplet. The second possibility is, in fact, the correct one, as shown by the completely deciphered genetic code in Figure 6–50. Each group of three consecutive nucleotides in RNA is called a **codon**, and each codon specifies either one amino acid or a stop to the translation process.

This genetic code is used universally in all present-day organisms. Although a few slight differences in the code have been found, these are chiefly in the DNA of mitochondria. Mitochondria have their own transcription and protein synthesis systems that operate quite independently from those of the rest of the cell, and it is understandable that their small genomes have been able to accommodate minor changes to the code (discussed in Chapter 14).

In principle, an RNA sequence can be translated in any one of three different **reading frames**, depending on where the decoding process begins (Figure 6–51). However, only one of the three possible reading frames in an mRNA encodes the required protein. We see later how a special punctuation signal at the beginning of each RNA message sets the correct reading frame at the start of protein synthesis.

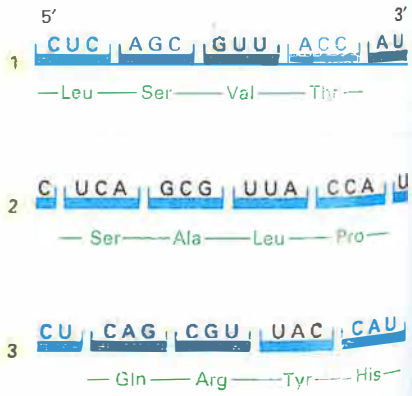


Figure 6–51 The three possible reading frames in protein synthesis. In the process of translating a nucleotide sequence *(blue)* into an amino acid sequence *(green)*, the sequence of nucleotides in an mRNA molecule is read from the 5′ to the 3′ end in sequential sets of three nucleotides. In principle, therefore, the same RNA sequence can specify three completely different amino acid sequences, depending on the reading frame. In reality, however, only one of these reading frames contains the actual message.

tRNA Molecules Match Amino Acids to Codons in mRNA

The codons in an mRNA molecule do not directly recognize the amino acids they specify: the group of three nucleotides does not, for example, bind directly to the amino acid. Rather, the translation of mRNA into protein depends on adaptor molecules that can recognize and bind both to the codon and, at another site on their surface, to the amino acid. These adaptors consist of a set of small RNA molecules known as **transfer RNAs (tRNAs)**, each about 80 nucleotides in length.

We saw earlier in this chapter that RNA molecules can fold up into precisely defined three-dimensional structures, and the tRNA molecules provide a striking example. Four short segments of the folded tRNA are double-helical, producing a molecule that looks like a cloverleaf when drawn schematically (Figure 6–52A). For example, a 5′-GCUC-3′ sequence in one part of a polynucleotide chain can form a relatively strong association with a 5′-GAGC-3′ sequence in another region of the same molecule. The cloverleaf undergoes further folding to form a compact L-shaped structure that is held together by additional hydrogen bonds between different regions of the molecule (Figure 6–52B,C).

Two regions of unpaired nucleotides situated at either end of the L-shaped molecule are crucial to the function of tRNA in protein synthesis. One of these regions forms the **anticodon,** a set of three consecutive nucleotides that pairs with the complementary codon in an mRNA molecule. The other is a short single-stranded region at the 3′ end of the molecule; this is the site where the amino acid that matches the codon is attached to the tRNA.

We have seen in the previous section that the genetic code is redundant; that is, several different codons can specify a single amino acid (see Figure 6–50). This redundancy implies either that there is more than one tRNA for many of the amino acids or that some tRNA molecules can base-pair with more than one

Figure 6–52 A tRNA molecule. In this series of diagrams, the same tRNA molecule—in this case a tRNA specific for the amino acid phenylalanine (Phe)—is depicted in various ways. (A) The cloverleaf structure, a convention used to show the complementary base-pairing *(red lines)* that creates the double-helical regions of the molecule. The anticodon is the sequence of three nucleotides that base-pairs with a codon in mRNA. The amino acid matching the codon/anticodon pair is attached at the 3′ end of the tRNA. tRNAs contain some unusual bases, which are produced by chemical modification after the tRNA has been synthesized. For example, the bases denoted ψ (for pseudouridine—see Figure 6–43) and D (for dihydrouridine—see Figure 6–55) are derived from uracil. (B and C) Views of the actual L-shaped molecule, based on x-ray diffraction analysis. Although a particular tRNA, that for the amino acid phenylalanine, is depicted, all other tRNAs have very similar structures. (D) The linear nucleotide sequence of the molecule, color-coded to match A, B, and C.

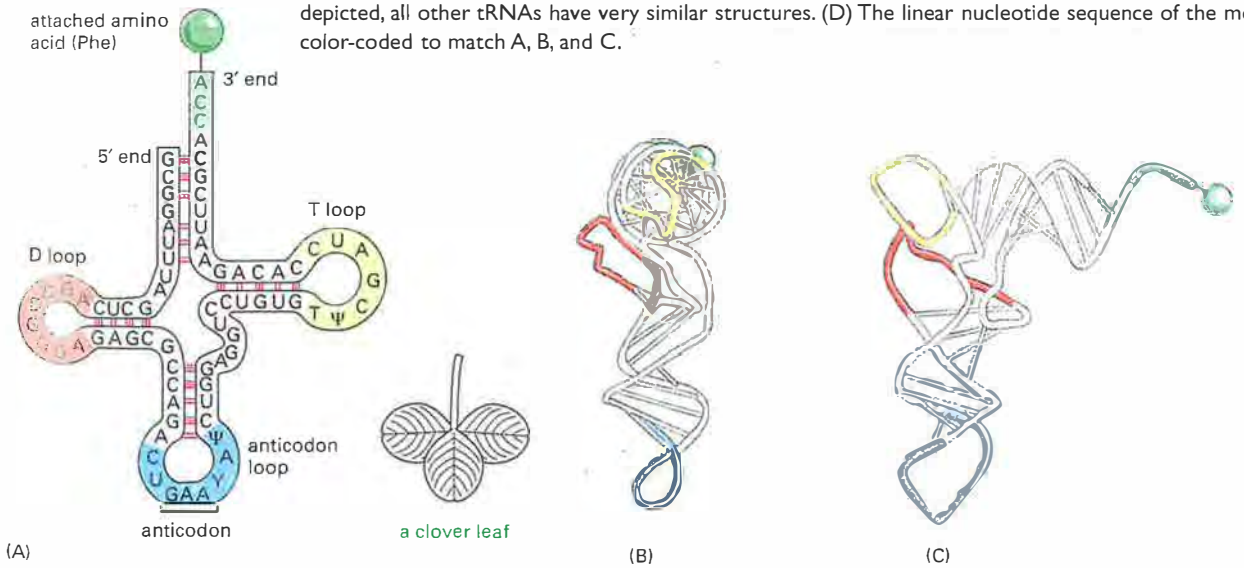
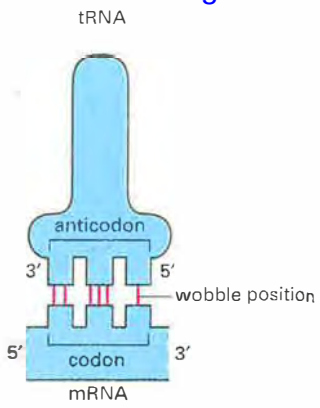


Figure 6–53 Wobble base-pairing between codons and anticodons. If the nucleotide listed in the first column is present at the third, or wobble, position of the codon, it can base-pair with any of the nucleotides listed in the second column. Thus, for example, when inosine (I) is present in the wobble position of the tRNA anticodon, the tRNA can recognize any one of three different codons in bacteria and either of two codons in eucaryotes. The inosine in tRNAs is formed from the deamination of guanine (see Figure 6–55), a chemical modification which takes place after the tRNA has been synthesized. The nonstandard base pairs, including those made with inosine, are generally weaker than conventional base pairs. Note that codon–anticodon base pairing is more stringent at positions 1 and 2 of the codon: here only conventional base pairs are permitted. The differences in wobble base-pairing interactions between bacteria and eucaryotes presumably result from subtle structural differences between bacterial and eucaryotic ribosomes, the molecular machines that perform protein synthesis. (Adapted from C. Guthrie and J. Abelson, in The Molecular Biology of the Yeast *Saccharomyces*: Metabolism and Gene Expression, pp. 487–528. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press, 1982.)



bacteria

wobble codon base	possible anticodon bases
U	A, G, or I
C	G or I
A	U or I
G	C or U

eucaryotes

wobble codon base	possible anticodon bases
U	G or I
C	G or I
A	U
G	C

codon. In fact, both situations occur. Some amino acids have more than one tRNA and some tRNAs are constructed so that they require accurate base-pairing only at the first two positions of the codon and can tolerate a mismatch (or *wobble*) at the third position (Figure 6–53). This wobble base-pairing explains why so many of the alternative codons for an amino acid differ only in their third nucleotide (see Figure 6–50). In bacteria, wobble base-pairings make it possible to fit the 20 amino acids to their 61 codons with as few as 31 kinds of tRNA molecules. The exact number of different kinds of tRNAs, however, differs from one species to the next. For example, humans have 497 tRNA genes but, among them, only 48 different anticodons are represented.

tRNAs Are Covalently Modified Before They Exit from the Nucleus

We have seen that most eucaryotic RNAs are covalently altered before they are allowed to exit from the nucleus, and tRNAs are no exception. Eucaryotic tRNAs are synthesized by RNA polymerase III. Both bacterial and eucaryotic tRNAs are typically synthesized as larger precursor tRNAs, and these are then trimmed to produce the mature tRNA. In addition, some tRNA precursors (from both bacteria and eucaryotes) contain introns that must be spliced out. This splicing reaction is chemically distinct from that of pre-mRNA splicing; rather than generating a lariat intermediate, tRNA splicing occurs through a cut-and-paste mechanism that is catalyzed by proteins (Figure 6–54). Trimming and splicing both require the precursor tRNA to be correctly folded in its cloverleaf configuration. Because misfolded tRNA precursors will not be processed properly, the

Figure 6–54 Structure of a tRNA-splicing endonuclease docked to a precursor tRNA. The endonuclease (a four-subunit enzyme) removes the tRNA intron *(blue)*. A second enzyme, a multifunctional tRNA ligase (not shown), then joins the two tRNA halves together. (Courtesy of Hong Li, Christopher Trotta, and John Abelson.)

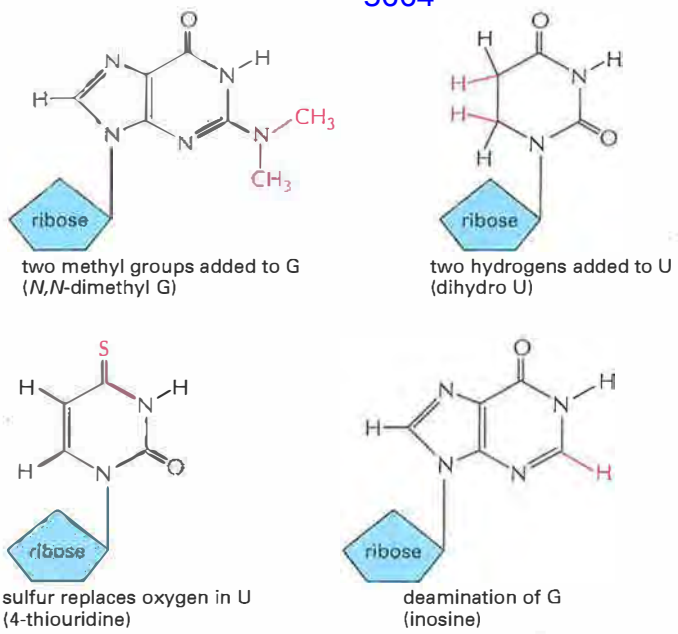


Figure 6–55 A few of the unusual nucleotides found in tRNA molecules. These nucleotides are produced by covalent modification of a normal nucleotide after it has been incorporated into an RNA chain. In most tRNA molecules about 10% of the nucleotides are modified (see Figure 6–52).

trimming and splicing reactions are thought to act as quality-control steps in the generation of tRNAs.

All tRNAs are also subject to a variety of chemical modifications—nearly one in 10 nucleotides in each mature tRNA molecule is an altered version of a standard G, U, C, or A ribonucleotide. Over 50 different types of tRNA modifications are known; a few are shown in Figure 6–55. Some of the modified nucleotides—most notably inosine, produced by the deamination of guanosine—affect the conformation and base-pairing of the anticodon and thereby facilitate the recognition of the appropriate mRNA codon by the tRNA molecule (see Figure 6–53). Others affect the accuracy with which the tRNA is attached to the correct amino acid.

Specific Enzymes Couple Each Amino Acid to Its Appropriate tRNA Molecule

We have seen that, to read the genetic code in DNA, cells make a series of different tRNAs. We now consider how each tRNA molecule becomes linked to the one amino acid in 20 that is its appropriate partner. Recognition and attachment of the correct amino acid depends on enzymes called **aminoacyl-tRNA synthetases**, which covalently couple each amino acid to its appropriate set of tRNA molecules (Figures 6–56 and 6–57). For most cells there is a different synthetase enzyme for each amino acid (that is, 20 synthetases in all); one attaches glycine

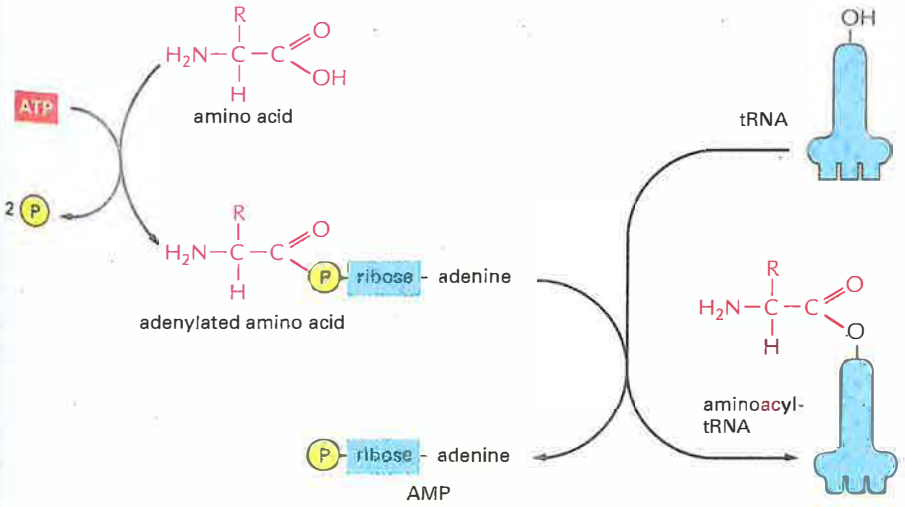


Figure 6–56 Amino acid activation. The two-step process in which an amino acid (with its side chain denoted by R) is activated for protein synthesis by an aminoacyl-tRNA synthetase enzyme is shown. As indicated, the energy of ATP hydrolysis is used to attach each amino acid to its tRNA molecule in a high-energy linkage. The amino acid is first activated through the linkage of its carboxyl group directly to an AMP moiety, forming an *adenylated amino acid*; the linkage of the AMP, normally an unfavorable reaction, is driven by the hydrolysis of the ATP molecule that donates the AMP. Without leaving the synthetase enzyme, the AMP-linked carboxyl group on the amino acid is then transferred to a hydroxyl group on the sugar at the 3′ end of the tRNA molecule. This transfer joins the amino acid by an activated ester linkage to the tRNA and forms the final aminoacyl-tRNA molecule. The synthetase enzyme is not shown in this diagram.

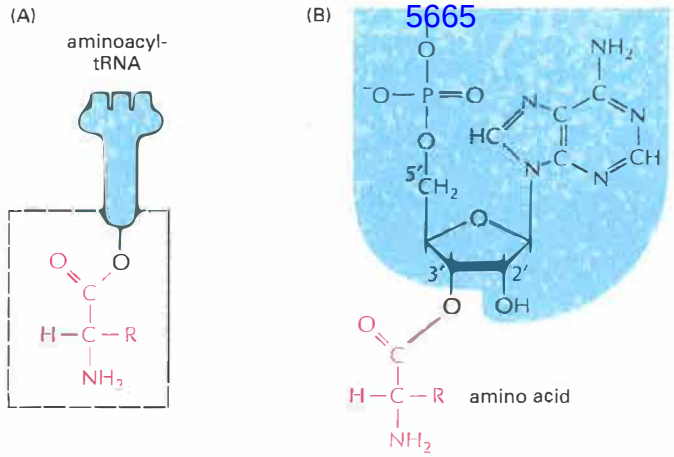


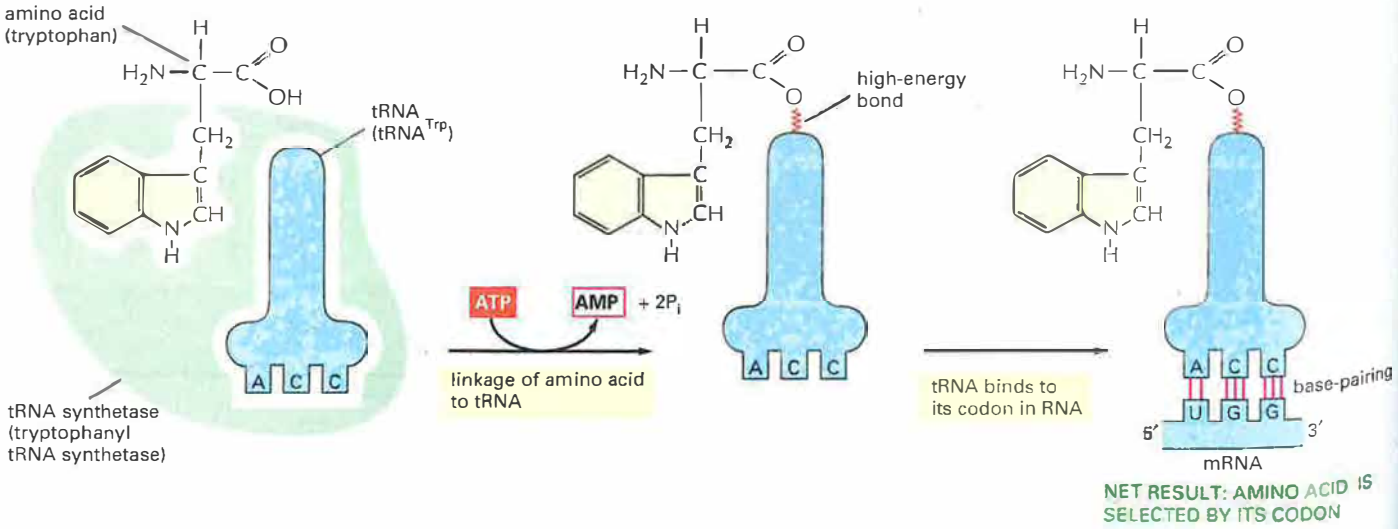
Figure 6–57 The structure of the aminoacyl-tRNA linkage. The carboxyl end of the amino acid forms an ester bond to ribose. Because the hydrolysis of this ester bond is associated with a large favorable change in free energy, an amino acid held in this way is said to be activated. (A) Schematic drawing of the structure. The amino acid is linked to the nucleotide at the 3′ end of the tRNA (see Figure 6–52). (B) Actual structure corresponding to boxed region in (A). These are two major classes of synthetase enzymes: one links the amino acid directly to the 3′-OH group of the ribose, and the other links it initially to the 2′-OH group. In the latter case, a subsequent transesterification reaction shifts the amino acid to the 3′ position. As in Figure 6–56, the "R-group" indicates the side chain of the amino acid.

to all tRNAs that recognize codons for glycine, another attaches alanine to all tRNAs that recognize codons for alanine, and so on. Many bacteria, however, have fewer than 20 synthetases, and the same synthetase enzyme is responsible for coupling more than one amino acid to the appropriate tRNAs. In these cases, a single synthetase places the identical amino acid on two different types of tRNAs, only one of which has an anticodon that matches the amino acid. A second enzyme then chemically modifies each "incorrectly" attached amino acid so that it now corresponds to the anticodon displayed by its covalently linked tRNA.

The synthetase-catalyzed reaction that attaches the amino acid to the 3′ end of the tRNA is one of many cellular reactions coupled to the energy-releasing hydrolysis of ATP (see pp. 83–84), and it produces a high-energy bond between the tRNA and the amino acid. The energy of this bond is used at a later stage in protein synthesis to link the amino acid covalently to the growing polypeptide chain.

Although the tRNA molecules serve as the final adaptors in converting nucleotide sequences into amino acid sequences, the aminoacyl-tRNA synthetase enzymes are adaptors of equal importance in the decoding process (Figure 6–58). This was established by an ingenious experiment in which an amino acid (cysteine) was chemically converted into a different amino acid (alanine) after it already had been attached to its specific tRNA. When such "hybrid" aminoacyl-tRNA molecules were used for protein synthesis in a cell-free system, the wrong amino acid was inserted at every point in the protein chain where that tRNA was used. Although cells have several quality control mechanisms to avoid this type of mishap, the experiment clearly establishes that the genetic code is translated by two sets of adaptors that act sequentially. Each matches one molecular surface to another with great specificity, and it is their combined

Figure 6–58 The genetic code is translated by means of two adaptors that act one after another. The first adaptor is the aminoacyl-tRNA synthetase, which couples a particular amino acid to its corresponding tRNA; the second adaptor is the tRNA molecule itself, whose *anticodon* forms base pairs with the appropriate *codon* on the mRNA. An error in either step would cause the wrong amino acid to be incorporated into a protein chain. In the sequence of events shown, the amino acid tryptophan (Trp) is selected by the codon UGG on the mRNA.



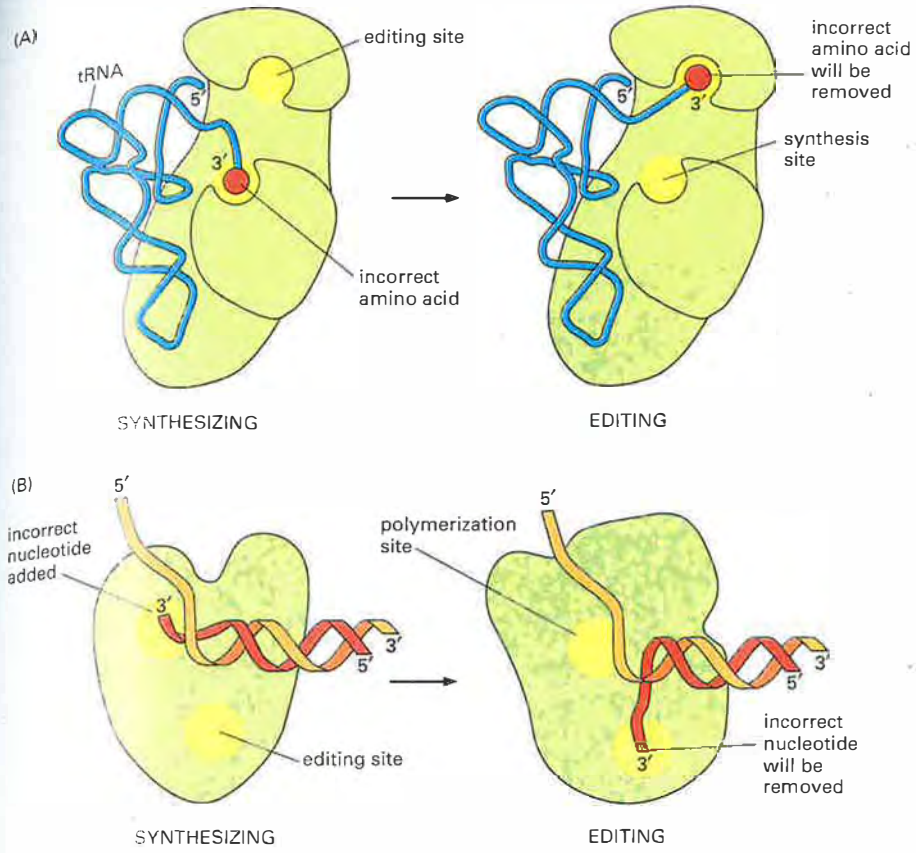


Figure 6–59 Hydrolytic editing. (A) tRNA synthetases remove their own coupling errors through hydrolytic editing of incorrectly attached amino acids. As described in the text, the correct amino acid is rejected by the editing site. (B) The error-correction process performed by DNA polymerase shows some similarities; however, it differs so far as the removal process depends strongly on a mispairing with the template (see Figure 5–9).

action that associates each sequence of three nucleotides in the mRNA molecule—that is, each codon—with its particular amino acid.

Editing by RNA Synthetases Ensures Accuracy

Several mechanisms working together ensure that the tRNA synthetase links the correct amino acid to each tRNA. The synthetase must first select the correct amino acid, and most do so by a two-step mechanism. First, the correct amino acid has the highest affinity for the active-site pocket of its synthetase and is therefore favored over the other 19. In particular, amino acids larger than the correct one are effectively excluded from the active site. However, accurate discrimination between two similar amino acids, such as isoleucine and valine (which differ by only a methyl group), is very difficult to achieve by a one-step recognition mechanism. A second discrimination step occurs after the amino acid has been covalently linked to AMP (see Figure 6–56). When tRNA binds the synthetase, it forces the amino acid into a second pocket in the synthetase, the precise dimensions of which exclude the correct amino acid but allow access by closely related amino acids. Once an amino acid enters this editing pocket, it is hydrolyzed from the AMP (or from the tRNA itself if the aminoacyl-tRNA bond has already formed) and released from the enzyme. This hydrolytic editing, which is analogous to the editing by DNA polymerases (Figure 6–59), raises the overall accuracy of tRNA charging to approximately one mistake in 40,000 couplings.

The tRNA synthetase must also recognize the correct set of tRNAs, and extensive structural and chemical complementarity between the synthetase and the tRNA allows various features of the tRNA to be sensed (Figure 6–60). Most tRNA synthetases directly recognize the matching tRNA anticodon; these synthetases contain three adjacent nucleotide-binding pockets, each of which is complementary in shape and charge to the nucleotide in the anticodon. For other synthetases it is the nucleotide sequence of the acceptor stem that is the key recognition determinant. In most cases, however, nucleotides at several positions on the tRNA are "read" by the synthetase.

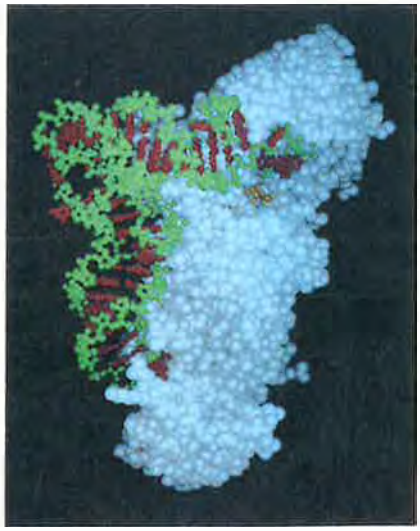


Figure 6–60 The recognition of a tRNA molecule by its aminoacyl-tRNA synthetase. For this tRNA ($tRNA^{Gln}$), specific nucleotides in both the anticodon *(bottom)* and the amino acid-accepting arm allow the correct tRNA to be recognized by the synthetase enzyme *(blue)*. (Courtesy of Tom Steitz.)

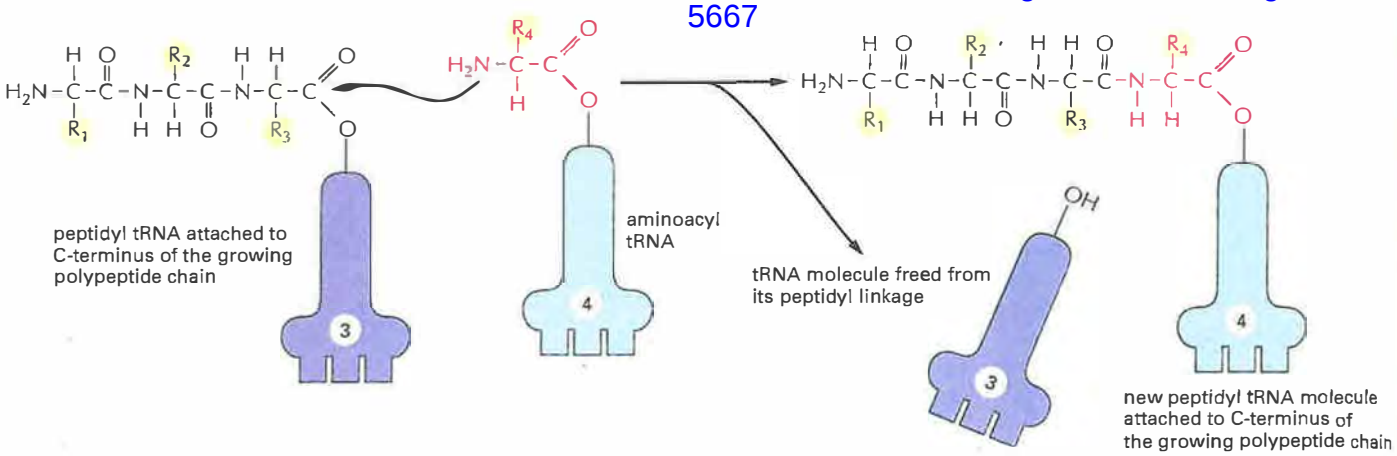


Figure 6–61 The incorporation of an amino acid into a protein. A polypeptide chain grows by the stepwise addition of amino acids to its C-terminal end. The formation of each peptide bond is energetically favorable because the growing C-terminus has been activated by the covalent attachment of a tRNA molecule. The peptidyl-tRNA linkage that activates the growing end is regenerated during each addition. The amino acid side chains have been abbreviated as R_1, R_2, R_3, and R_4; as a reference point, all of the atoms in the second amino acid in the polypeptide chain are shaded *gray*. The figure shows the addition of the fourth amino acid to the growing chain.

Amino Acids Are Added to the C-terminal End of a Growing Polypeptide Chain

Having seen that amino acids are first coupled to tRNA molecules, we now turn to the mechanism by which they are joined together to form proteins. The fundamental reaction of protein synthesis is the formation of a peptide bond between the carboxyl group at the end of a growing polypeptide chain and a free amino group on an incoming amino acid. Consequently, a protein is synthesized stepwise from its N-terminal end to its C-terminal end. Throughout the entire process the growing carboxyl end of the polypeptide chain remains activated by its covalent attachment to a tRNA molecule (a peptidyl-tRNA molecule). This high-energy covalent linkage is disrupted during each addition but is immediately replaced by the identical linkage on the most recently added amino acid (Figure 6–61). In this way, each amino acid added carries with it the activation energy for the addition of the next amino acid rather than the energy for its own addition—an example of the "head growth" type of polymerization described in Figure 2–68.

The RNA Message Is Decoded on Ribosomes

As we have seen, the synthesis of proteins is guided by information carried by mRNA molecules. To maintain the correct reading frame and to ensure accuracy (about 1 mistake every 10,000 amino acids), protein synthesis is performed in the **ribosome**, a complex catalytic machine made from more than 50 different proteins (the *ribosomal proteins*) and several RNA molecules, the **ribosomal RNAs (rRNAs)**. A typical eucaryotic cell contains millions of ribosomes in its cytoplasm (Figure 6–62). As we have seen, eucaryotic ribosomal subunits are

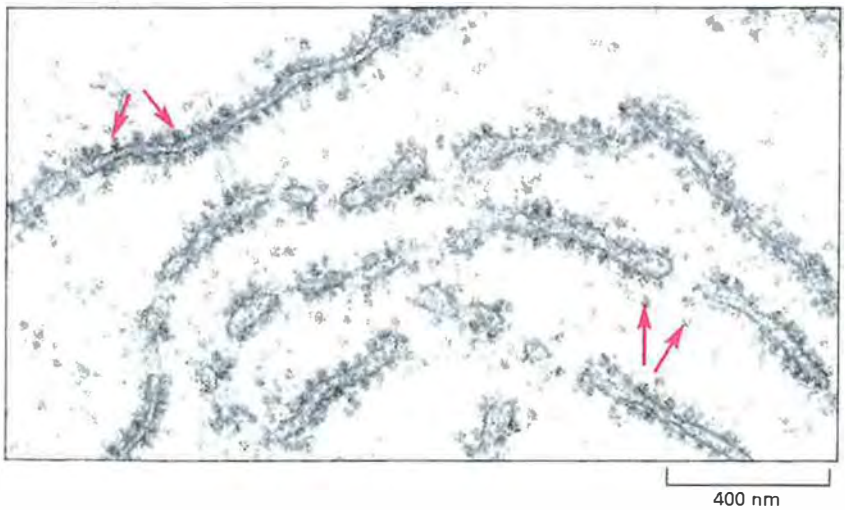


Figure 6–62 Ribosomes in the cytoplasm of a eucaryotic cell. This electron micrograph shows a thin section of a small region of cytoplasm. The ribosomes appear as black dots *(red arrows)*. Some are free in the cytosol; others are attached to membranes of the endoplasmic reticulum. (Courtesy of Daniel S. Friend.)

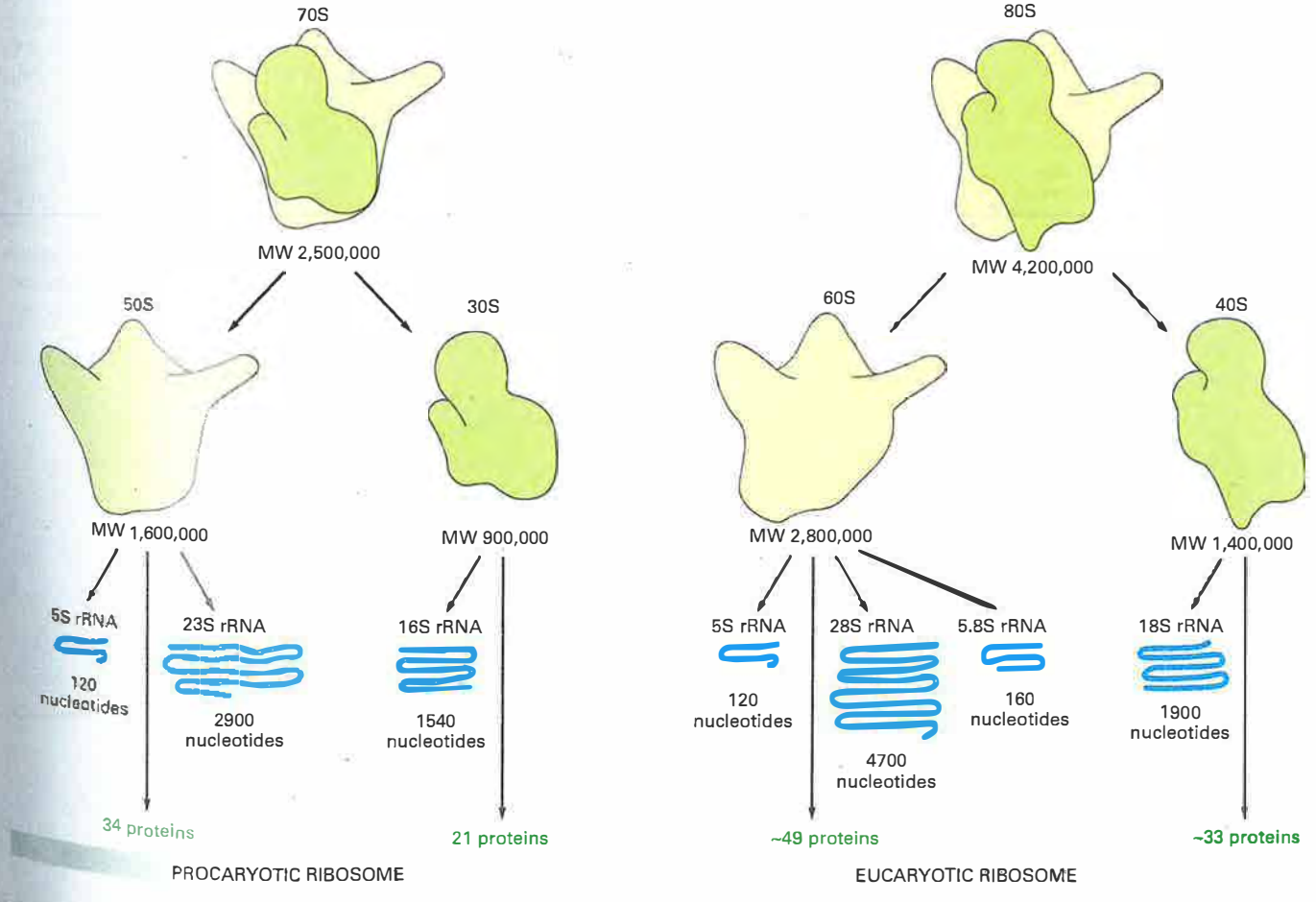
assembled at the nucleolus, by the association of newly transcribed and modified rRNAs with ribosomal proteins, which have been transported into the nucleus after their synthesis in the cytoplasm. The two ribosomal subunits are then exported to the cytoplasm, where they perform protein synthesis.

Eucaryotic and procaryotic ribosomes are very similar in design and function. Both are composed of one large and one small subunit that fit together to form a complete ribosome with a mass of several million daltons (Figure 6–63). The small subunit provides a framework on which the tRNAs can be accurately matched to the codons of the mRNA (see Figure 6–58), while the large subunit catalyzes the formation of the peptide bonds that link the amino acids together into a polypeptide chain (see Figure 6–61).

When not actively synthesizing proteins, the two subunits of the ribosome are separate. They join together on an mRNA molecule, usually near its 5′ end, to initiate the synthesis of a protein. The mRNA is then pulled through the ribosome; as its codons encounter the ribosome's active site, the mRNA nucleotide sequence is translated into an amino acid sequence using the tRNAs as adaptors to add each amino acid in the correct sequence to the end of the growing polypeptide chain. When a stop codon is encountered, the ribosome releases the finished protein, its two subunits separate again. These subunits can then be used to start the synthesis of another protein on another mRNA molecule.

Ribosomes operate with remarkable efficiency: in one second, a single ribosome of a eucaryotic cell adds about 2 amino acids to a polypeptide chain; the ribosomes of bacterial cells operate even faster, at a rate of about 20 amino acids

Figure 6–63 A comparison of the structures of procaryotic and eucaryotic ribosomes. Ribosomal components are commonly designated by their "S values," which refer to their rate of sedimentation in an ultracentrifuge. Despite the differences in the number and size of their rRNA and protein components, both procaryotic and eucaryotic ribosomes have nearly the same structure and they function similarly. Although the 18S and 28S rRNAs of the eucaryotic ribosome contain many extra nucleotides not present in their bacterial counterparts, these nucleotides are present as multiple insertions that form extra domains and leave the basic structure of each rRNA largely unchanged.



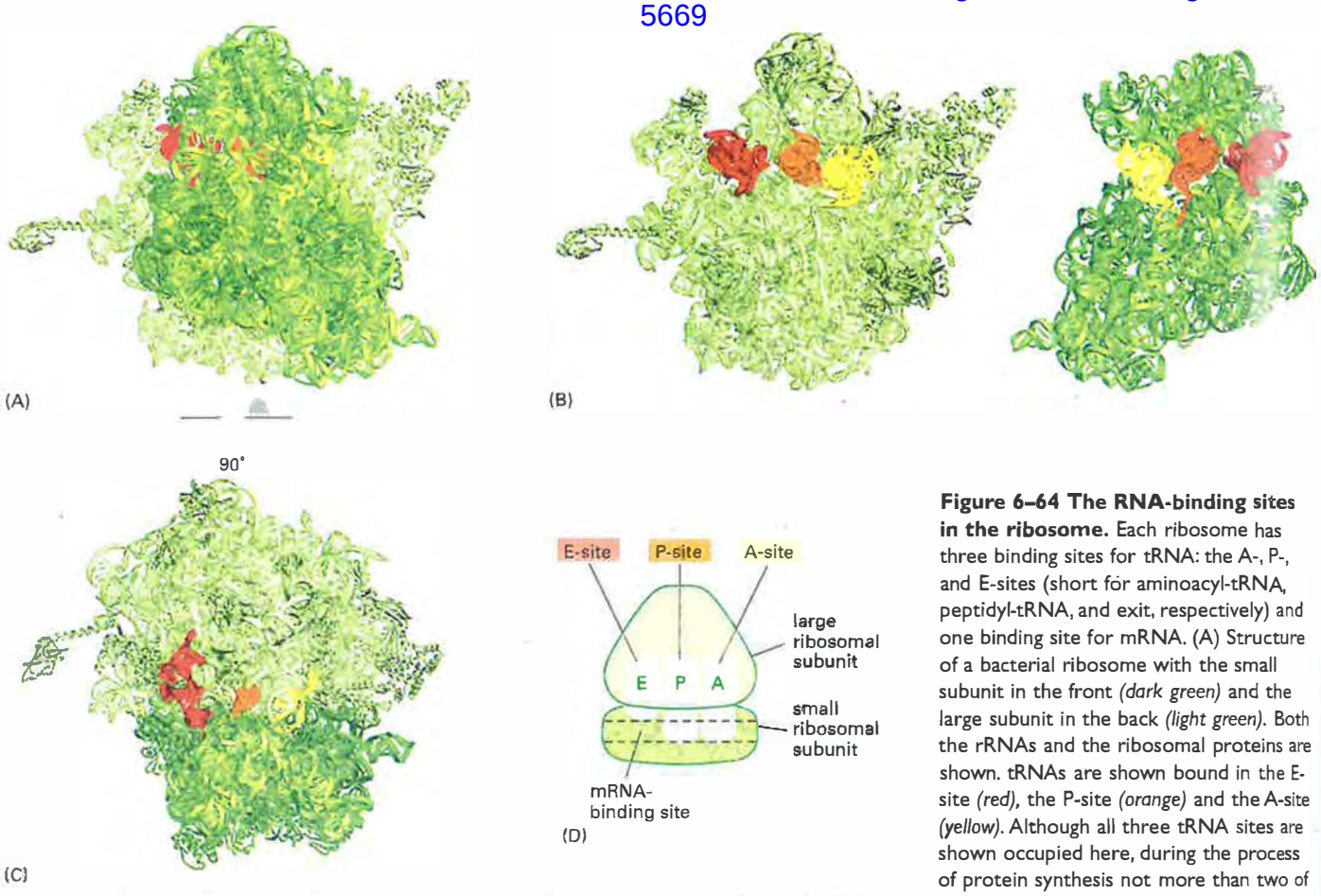


Figure 6–64 The RNA-binding sites in the ribosome. Each ribosome has three binding sites for tRNA: the A-, P-, and E-sites (short for aminoacyl-tRNA, peptidyl-tRNA, and exit, respectively) and one binding site for mRNA. (A) Structure of a bacterial ribosome with the small subunit in the front *(dark green)* and the large subunit in the back *(light green)*. Both the rRNAs and the ribosomal proteins are shown. tRNAs are shown bound in the E-site *(red)*, the P-site *(orange)* and the A-site *(yellow)*. Although all three tRNA sites are shown occupied here, during the process of protein synthesis not more than two of these sites are thought to contain tRNA molecules at any one time (see Figure 6–65). (B) Structure of the large *(left)* and small *(right)* ribosomal subunits arranged as though the ribosome in (A) were opened like a book. (C) Structure of the ribosome in (A) viewed from the top. (D) Highly schematic representation of a ribosome (in the same orientation as C), which will be used in subsequent figures. (A, B, and C, adapted from M.M. Yusupov et al., *Science* 292:883–896, 2001, courtesy of Albion Bausom and Harry Noller.)

per second. How does the ribosome choreograph the many coordinated movements required for efficient translation? A ribosome contains four binding sites for RNA molecules: one is for the mRNA and three (called the A-site, the P-site, and the E-site) are for tRNAs (Figure 6–64). A tRNA molecule is held tightly at the A- and P-sites only if its anticodon forms base pairs with a complementary codon (allowing for wobble) on the mRNA molecule that is bound to the ribosome. The A- and P-sites are close enough together for their two tRNA molecules to be forced to form base pairs with adjacent codons on the mRNA molecule. This feature of the ribosome maintains the correct reading frame on the mRNA.

Once protein synthesis has been initiated, each new amino acid is added to the elongating chain in a cycle of reactions containing three major steps. Our description of the chain elongation process begins at a point at which some amino acids have already been linked together and there is a tRNA molecule in the P-site on the ribosome, covalently joined to the end of the growing polypeptide (Figure 6–65). In step 1, a tRNA carrying the next amino acid in the chain binds to the ribosomal A-site by forming base pairs with the codon in mRNA positioned there, so that the P-site and the A-site contain adjacent bound tRNAs. In step 2, the carboxyl end of the polypeptide chain is released from the tRNA at the P-site (by breakage of the high-energy bond between the tRNA and its amino acid) and joined to the free amino group of the amino acid linked to the tRNA at the A-site, forming a new peptide bond. This central reaction of protein synthesis is catalyzed by a *peptidyl transferase* catalytic activity contained in the large ribosomal subunit. This reaction is accompanied by several conformational changes in the ribosome, which shift the two tRNAs into the E- and P-sites of the large subunit. In step 3, another series of conformational changes moves the mRNA exactly three nucleotides through the ribosome and resets the ribosome so it is ready to receive the next amino acyl tRNA. Step 1 is then repeated with a new incoming aminoacyl tRNA, and so on.

This three-step cycle is repeated each time an amino acid is added to the polypeptide chain, and the chain grows from its amino to its carboxyl end until a stop codon is encountered.

Elongation Factors Drive Translation Forward

The basic cycle of polypeptide elongation shown in outline in Figure 6–65 has an additional feature that makes translation especially efficient and accurate. Two *elongation factors* (EF-Tu and EF-G) enter and leave the ribosome during each cycle, each hydrolyzing GTP to GDP and undergoing conformational changes in the process. Under some conditions, ribosomes can be made to perform protein synthesis without the aid of the elongation factors and GTP hydrolysis, but this synthesis is very slow, inefficient, and inaccurate. The process is speeded up enormously by coupling conformational changes in the elongation factors to transitions between different conformational states of the ribosome. Although these conformational changes in the ribosome are not yet understood in detail, some may involve RNA rearrangements similar to those occurring in the RNAs of the spliceosome (see Figure 6–30). The cycles of elongation factor association, GTP hydrolysis, and dissociation ensures that the conformational changes occur in the "forward" direction and translation thereby proceeds efficiently (Figure 6–66).

In addition to helping move translation forward, EF-Tu is thought to increase the accuracy of translation by monitoring the initial interaction between a charged tRNA and a codon (see Figure 6–66). Charged tRNAs enter the ribosome bound to the GTP-form of EF-Tu. Although the bound elongation factor allows codon–anticodon pairing to occur, it prevents the amino acid from being incorporated into the growing polypeptide chain. The initial codon recognition, however, triggers the elongation factor to hydrolyze its bound GTP (to GDP and inorganic phosphate), whereupon the factor dissociates from the ribosome without its tRNA, allowing protein synthesis to proceed. The elongation factor introduces two short delays between codon–anticodon base pairing and polypeptide chain elongation; these delays selectively permit incorrectly bound tRNAs to exit from the ribosome before the irreversible step of chain elongation occurs. The first delay is the time required for GTP hydrolysis. The rate of GTP hydrolysis by EF-Tu is faster for a correct codon–anticodon pair than for an incorrect pair; hence an incorrectly bound tRNA molecule has a longer window of opportunity to dissociate from the ribosome. In other words, GTP hydrolysis selectively captures the correctly bound tRNAs. A second lag occurs between EF-Tu dissociation and the full accommodation of the tRNA in the A site of the ribosome. Although this lag is believed to be the same for correctly and incorrectly bound tRNAs, an incorrect tRNA molecule forms a smaller number of codon–anticodon hydrogen bonds than does a correctly matched pair and is therefore more likely to dissociate during this period. These two delays introduced by the elongation factor cause most incorrectly bound tRNA molecules (as well as a significant number of correctly bound molecules) to leave the

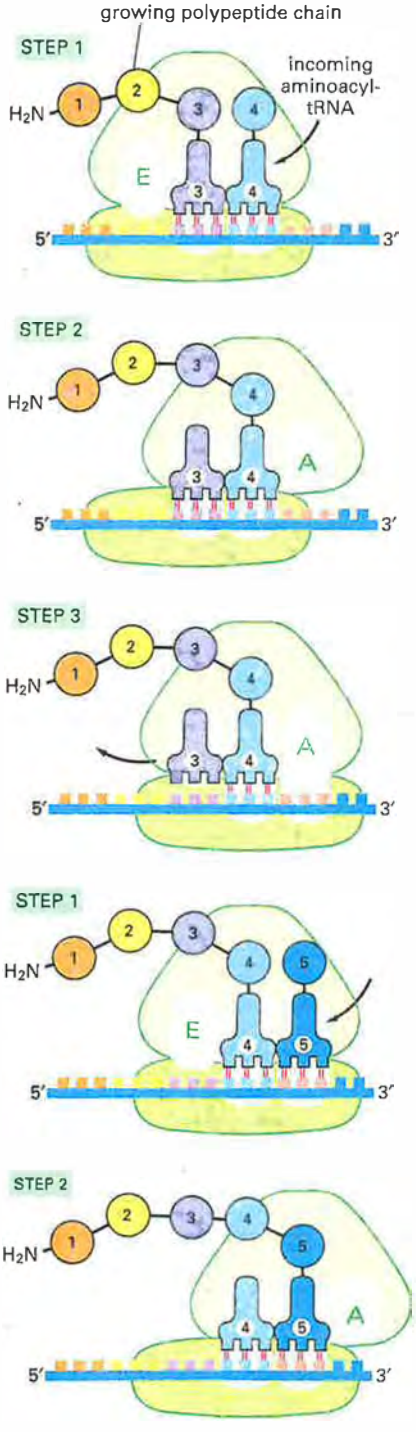


Figure 6–65 Translating an mRNA molecule. Each amino acid added to the growing end of a polypeptide chain is selected by complementary base-pairing between the anticodon on its attached tRNA molecule and the next codon on the mRNA chain. Because only one of the many types of tRNA molecules in a cell can base-pair with each codon, the codon determines the specific amino acid to be added to the growing polypeptide chain. The three-step cycle shown is repeated over and over during the synthesis of a protein. An aminoacyl-tRNA molecule binds to a vacant A-site on the ribosome in step 1, a new peptide bond is formed in step 2, and the mRNA moves a distance of three nucleotides through the small-subunit chain in step 3, ejecting the spent tRNA molecule and "resetting" the ribosome so that the next aminoacyl-tRNA molecule can bind. Although the figure shows a large movement of the small ribosome subunit relative to the large subunit, the conformational changes that actually take place in the ribosome during translation are more subtle. It is likely that they involve a series of small rearrangements within each subunit as well as several small shifts between the two subunits. As indicated, the mRNA is translated in the 5′-to-3′ direction, and the N-terminal end of a protein is made first, with each cycle adding one amino acid to the C-terminus of the polypeptide chain. The position at which the growing peptide chain is attached to a tRNA does not change during the elongation cycle: it is always linked to the tRNA present in the P site of the large subunit.

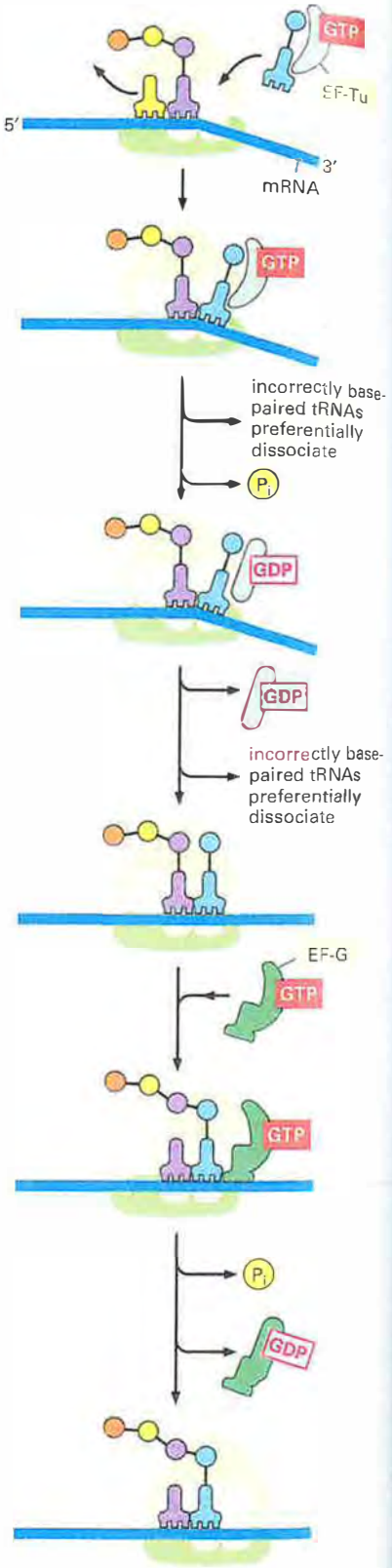


Figure 6–66 Detailed view of the translation cycle. The outline of translation presented in Figure 6–65 has been supplemented with additional features, including the participation of elongation factors and a mechanism by which translational accuracy is improved. In the initial binding event *(top panel)* an aminoacyl-tRNA molecule that is tightly bound to EF-Tu pairs transiently with the codon at the A-site in the small subunit. During this step *(second panel)*, the tRNA occupies a hybrid-binding site on the ribosome. The codon–anticodon pairing triggers GTP hydrolysis by EF-Tu causing it to dissociate from the aminoacyl-tRNA, which now enters the A-site *(fourth panel)* and can participate in chain elongation. A delay between aminoacyl-tRNA binding and its availability for protein synthesis is thereby inserted into the protein synthesis mechanism. As described in the text, this delay increases the accuracy of translation. In subsequent steps, elongation factor EF-G in the GTP-bound form enters the ribosome and binds in or near the A-site on the large ribosomal subunit, accelerating the movement of the two bound tRNAs into the A/P and P/E hybrid states. Contact with the ribosome stimulates the GTPase activity of EF-G, causing a dramatic conformational change in EF-G as it switches from the GTP to the GDP-bound form. This change moves the tRNA bound to the A/P hybrid state to the P-site and advances the cycle of translation forward by one codon.

During each cycle of translation elongation, the tRNAs molecules move through the ribosome in an elaborate series of gyrations during which they transiently occupy several "hybrid" binding states. In one, the tRNA is simultaneously bound to the A site of the small subunit and the P site of the large subunit; in another, the tRNA is bound to the P site of the small subunit and the E site of the large subunit. In a single cycle, a tRNA molecule is considered to occupy six different sites, the initial binding site (called the A/T hybrid state), the A/A site, the A/P hybrid state, the P/P site, the P/E hybrid state, and the E-site. Each tRNA is thought to ratchet through these positions, undergoing rotations along its long axis at each change in location.

EF-Tu and EF-G are the designations used for the bacterial elongation factors; in eucaryotes, they are called EF-1 and EF-2, respectively. The dramatic change in the three-dimensional structure of EF-Tu that is caused by GTP hydrolysis was illustrated in Figure 3–74. For each peptide bond formed, a molecule of EF-Tu and EF-G are each released in their inactive, GDP-bound forms. To be used again, these proteins must have their GDP exchanged for GTP. In the case of EF-Tu, this exchange is performed by a specific member of a large class of proteins known as *GTP exchange factors*.

ribosome without being used for protein synthesis, and this two-step mechanism is largely responsible for the 99.99% accuracy of the ribosome in translating proteins.

Recent discoveries indicate that EF-Tu may have an additional role in raising the overall accuracy of translation. Earlier in this chapter, we discussed the key role of aminoacyl synthetases in accurately matching amino acids to tRNAs. As the GTP-bound form of EF-Tu escorts aminoacyl-tRNAs to the ribosome (see Figure 6–66), it apparently double-checks for the proper correspondence between amino acid and tRNA and rejects those that are mismatched. Exactly how this is accomplished is not well-understood, but it may involve the overall binding energy between EF-Tu and the aminoacyl-tRNA. According to this idea, correct matches have a narrowly defined affinity for EF-Tu, and incorrect matches bind either too strongly or too weakly. EF-Tu thus appears to discriminate, albeit crudely, among many different amino acid-tRNA combinations, selectively allowing only the correct ones to enter the ribosome.

The Ribosome Is a Ribozyme

The ribosome is a very large and complex structure, composed of two-thirds RNA and one-third protein. The determination, in 2000, of the entire three-dimensional structure of its large and small subunits is a major triumph of modern structural biology. The structure strongly confirms the earlier evidence that rRNAs—and not proteins—are responsible for the ribosome's overall structure, its ability to position tRNAs on the mRNA, and its catalytic activity in forming covalent peptide bonds. Thus, for example, the ribosomal RNAs are folded into

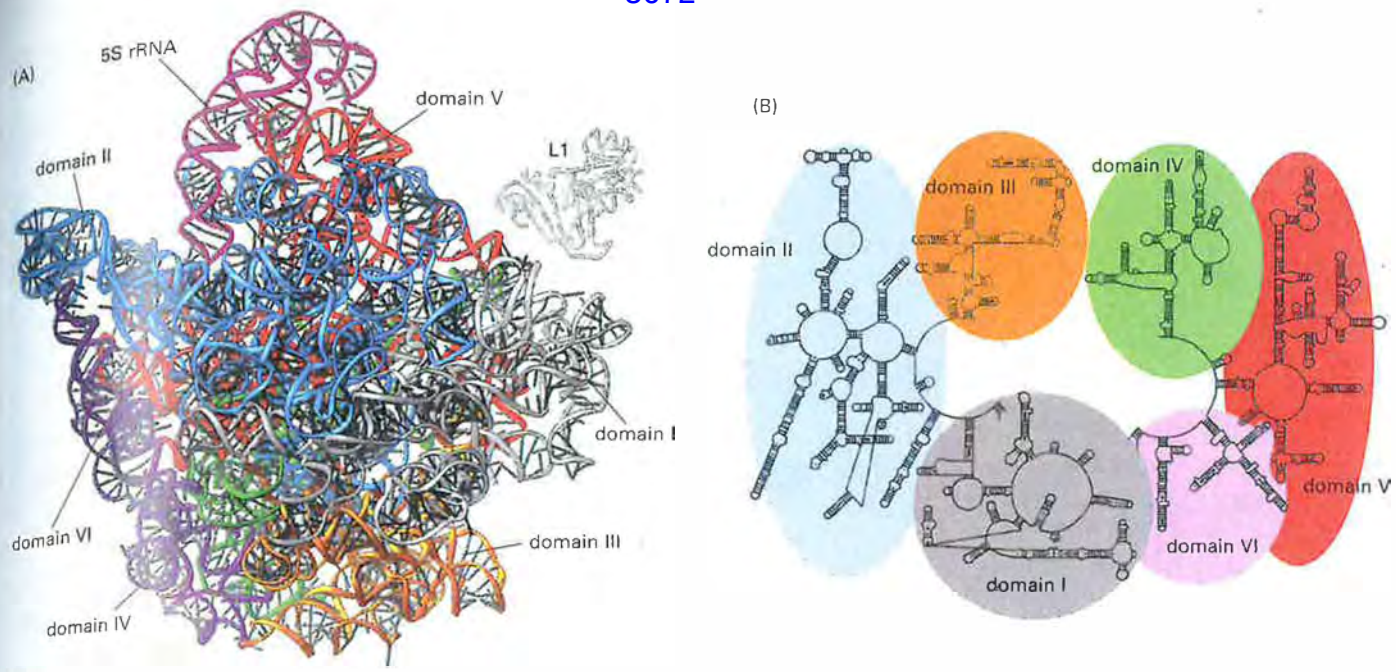


Figure 6–67 Structure of the rRNAs in the large subunit of a bacterial ribosome, as determined by x-ray crystallography. (A) Three-dimensional structures of the large-subunit rRNAs (5S and 23S) as they appear in the ribosome. One of the protein subunits of the ribosome (L1) is also shown as a reference point, since it forms a characteristic protrusion on the ribosome. (B) Schematic diagram of the secondary structure of the 23S rRNA showing the extensive network of base-pairing. The structure has been divided into six structural 'domains' whose colors correspond to those of the three-dimensional structure in (A). The secondary-structure diagram is highly schematized to represent as much of the structure as possible in two dimensions. To do this, several discontinuities in the RNA chain have been introduced, although in reality the 23S RNA is a single RNA molecule. For example, the base of Domain III is continuous with the base of Domain IV even though a gap appears in the diagram. (Adapted from N. Ban et al., *Science* 289:905–920, 2000.)

highly compact, precise three-dimensional structures that form the compact core of the ribosome and thereby determine its overall shape (Figure 6–67).

In marked contrast to the central positions of the rRNA, the ribosomal proteins are generally located on the surface and fill in the gaps and crevices of the folded RNA (Figure 6–68). Some of these proteins contain globular domains on the ribosome surface that send out extended regions of polypeptide chain that penetrate short distances into holes in the RNA core (Figure 6–69). The main role

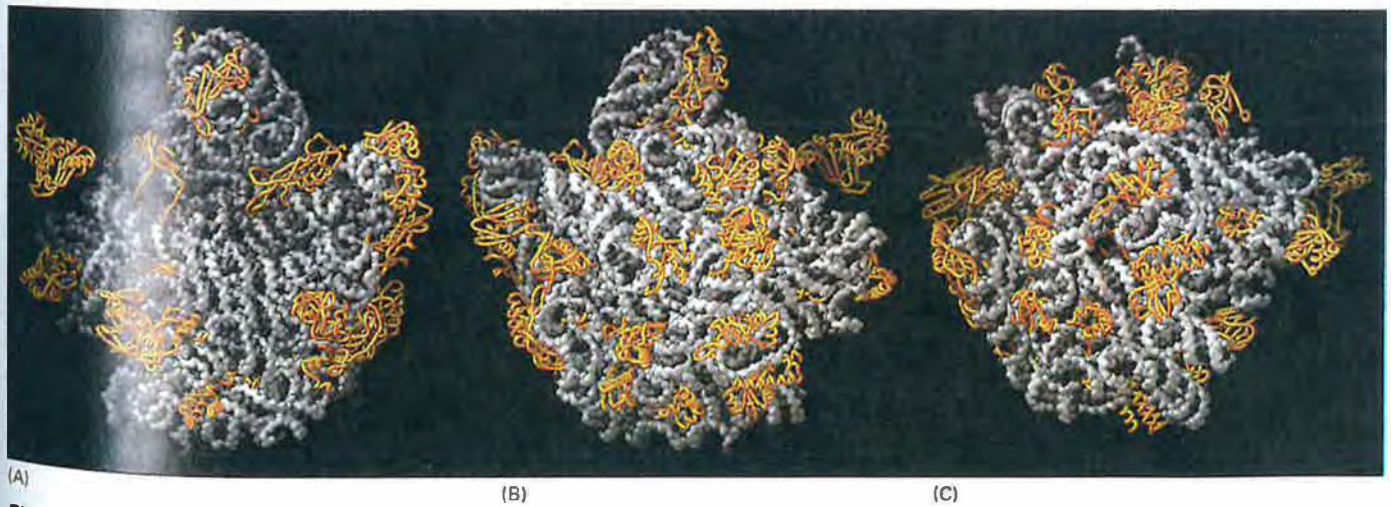


Figure 6–68 Location of the protein components of the bacterial large ribosomal subunit. The rRNAs (5S and 23S) are depicted in gray and the large-subunit proteins (27 of the 31 total) in *gold*. For convenience, the protein structures depict only the polypeptide backbones. (A) View of the interface with the small subunit, the same view shown in Figure 6–64B. (B) View of the back of the large subunit, obtained by rotating (A) by 180° around a vertical axis. (C) View of the bottom of the large subunit showing the peptide exit channel in the center of the structure. (From N. Ban et al., *Science* 289:905–920, 2000. © AAAS.)

of the ribosomal proteins seems to be to stabilize the RNA core, while permitting the changes in rRNA conformation that are necessary for this RNA to catalyze efficient protein synthesis.

Not only are the three binding sites for tRNAs (the A-, P-, and E-sites) on the ribosome formed primarily by the ribosomal RNAs, but the catalytic site for peptide bond formation is clearly formed by the 23S RNA, with the nearest amino acid located more than 1.8 nm away. This RNA-based catalytic site for peptidyl transferase is similar in many respects to those found in some proteins; it is a highly structured pocket that precisely orients the two reactants (the growing peptide chain and an aminoacyl-tRNA), and it provides a functional group to act as a general acid–base catalyst—in this case apparently, a ring nitrogen of adenine, instead of an amino acid side chain such as histidine (Figure 6–70). The ability of an RNA molecule to act as such a catalyst was initially surprising because RNA was thought to lack an appropriate chemical group that could both accept and donate a proton. Although the pK of adenine-ring nitrogens is usually around 3.5, the three-dimensional structure and charge distribution of the 23S rRNA active site force the pK of this apparently critical adenine into the neutral range and thereby create the enzymatic activity.

RNA molecules that possess catalytic activity are known as **ribozymes**. We saw earlier in this chapter how other ribozymes function in RNA-splicing reactions (for example, see Figure 6–36). In the final section of this chapter, we consider what the recently recognized ability of RNA molecules to function as catalysts for a wide variety of different reactions might mean for the early evolution of living cells. Here we need only note that there is good reason to suspect that RNA rather than protein molecules served as the first catalysts for living cells. If so, the ribosome, with its RNA core, might be viewed as a relic of an earlier time in life's history—when protein synthesis evolved in cells that were run almost entirely by ribozymes.

Figure 6–69 Structure of the L15 protein in the large subunit of the bacterial ribosome. The globular domain of the protein lies on the surface of the ribosome and an extended region penetrates deeply into the RNA core of the ribosome. The L15 protein is shown in *yellow* and a portion of the ribosomal RNA core is shown in *red*. (Courtesy of D. Klein, P.B. Moore and T.A. Steitz.)

Nucleotide Sequences in mRNA Signal Where to Start Protein Synthesis

The initiation and termination of translation occur through variations on the translation elongation cycle described above. The site at which protein synthesis begins on the mRNA is especially crucial, since it sets the reading frame for the whole length of the message. An error of one nucleotide either way at this stage would cause every subsequent codon in the message to be misread, so that a nonfunctional protein with a garbled sequence of amino acids would result. The initiation step is also of great importance in another respect, since for most genes it is the last point at which the cell can decide whether the mRNA is to be translated and the protein synthesized; the rate of initiation thus determines the rate at which the protein is synthesized. We shall see in Chapter 7 that cells use several mechanisms to regulate translation initiation.

The translation of an mRNA begins with the codon AUG, and a special tRNA is required to initiate translation. This **initiator tRNA** always carries the amino acid methionine (in bacteria, a modified form of methionine—formylmethionine—is used) so that all newly made proteins have methionine as the first amino acid at their N-terminal end, the end of a protein that is synthesized first. This

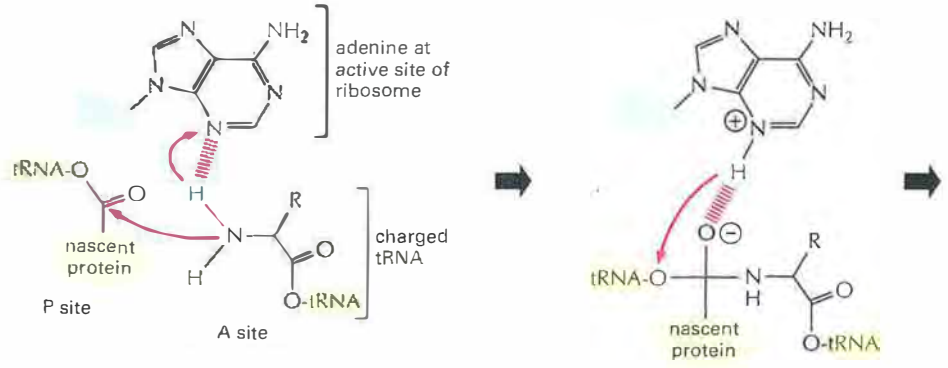


Figure 6–70 A possible reaction mechanism for the peptidyl transferase activity present in the large ribosomal subunit. The overall reaction is catalyzed by an active site in the 23S rRNA. In the first step of the proposed mechanism, the N3 of the active-site adenine abstracts a proton from the amino acid attached to the tRNA at the ribosome's A-site, allowing its amino nitrogen to attack the carboxyl group at the end of the growing peptide chain. In the next step this protonated adenine donates its hydrogen to the oxygen linked to the peptidyl-tRNA, causing this tRNA's release from the peptide chain. This leaves a polypeptide chain that is one amino acid longer than the starting reactants. The entire reaction cycle would then repeat with the next aminoacyl tRNA that enters the A-site. (Adapted from P. Nissen et al., *Science* 289:920–930, 2000.)

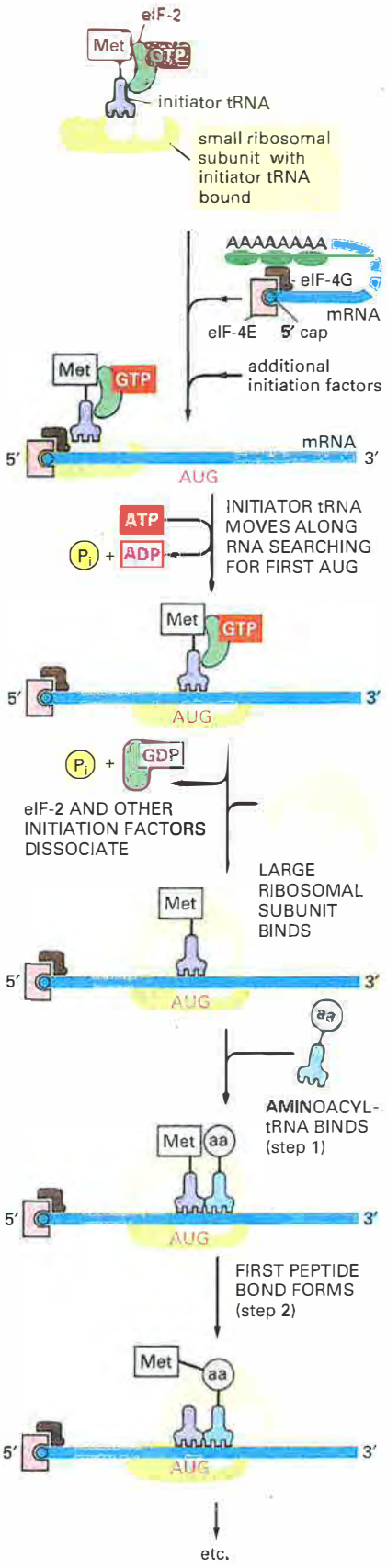


Figure 6–71 The initiation phase of protein synthesis in eucaryotes. Only three of the many translation initiation factors required for this process are shown. Efficient translation initiation also requires the poly-A tail of the mRNA bound by poly-A-binding proteins which, in turn, interact with eIF4G. In this way, the translation apparatus ascertains that both ends of the mRNA are intact before initiating (see Figure 6–40). Although only one GTP hydrolysis event is shown in the figure, a second is known to occur just before the large and small ribosomal subunits join.

methionine is usually removed later by a specific protease. The initiator tRNA has a nucleotide sequence distinct from that of the tRNA that normally carries methionine.

In eucaryotes, the initiator tRNA (which is coupled to methionine) is first loaded into the small ribosomal subunit along with additional proteins called **eucaryotic initiation factors**, or **eIFs** (Figure 6–71). Of all the aminoacyl tRNAs in the cell, only the methionine-charged initiator tRNA is capable of tightly binding the small ribosome subunit without the complete ribosome present. Next, the small ribosomal subunit binds to the 5′ end of an mRNA molecule, which is recognized by virtue of its 5′ cap and its two bound initiation factors, eIF4E (which directly binds the cap) and eIF4G (see Figure 6–40). The small ribosomal subunit then moves forward (5′ to 3′) along the mRNA, searching for the first AUG. This movement is facilitated by additional initiation factors that act as ATP-powered helicases, allowing the small subunit to scan through RNA secondary structure. In 90% of mRNAs, translation begins at the first AUG encountered by the small subunit. At this point, the initiation factors dissociate from the small ribosomal subunit to make way for the large ribosomal subunit to assemble with it and complete the ribosome. The initiator tRNA is now bound to the P-site, leaving the A-site vacant. Protein synthesis is therefore ready to begin with the addition of the next aminoacyl tRNA molecule (see Figure 6–71).

The nucleotides immediately surrounding the start site in eucaryotic mRNAs influence the efficiency of AUG recognition during the above scanning process. If this recognition site is quite different from the consensus recognition sequence, scanning ribosomal subunits will sometimes ignore the first AUG codon in the mRNA and skip to the second or third AUG codon instead. Cells frequently use this phenomenon, known as "leaky scanning," to produce two or more proteins, differing in their N-termini, from the same mRNA molecule. It allows some genes to produce the same protein with and without a signal sequence attached at its N-terminus, for example, so that the protein is directed to two different compartments in the cell.

The mechanism for selecting a start codon in bacteria is different. Bacterial mRNAs have no 5′ caps to tell the ribosome where to begin searching for the start of translation. Instead, each bacterial mRNA contains a specific ribosome-binding site (called the Shine–Dalgarno sequence, named after its discoverers) that is located a few nucleotides upstream of the AUG at which translation is to begin. This nucleotide sequence, with the consensus 5′-AGGAGGU-3′, forms base pairs with the 16S rRNA of the small ribosomal subunit to position the initiating AUG codon in the ribosome. A set of translation initiation factors orchestrates this interaction, as well as the subsequent assembly of the large ribosomal subunit to complete the ribosome.

Unlike a eucaryotic ribosome, a bacterial ribosome can therefore readily assemble directly on a start codon that lies in the interior of an mRNA molecule, so long as a ribosome-binding site precedes it by several nucleotides. As a result, bacterial mRNAs are often *polycistronic*—that is, they encode several different proteins, each of which is translated from the same mRNA molecule (Figure 6–72). In contrast, a eucaryotic mRNA generally encodes only a single protein.

Stop Codons Mark the End of Translation

The end of the protein-coding message is signaled by the presence of one of three codons (UAA, UAG, or UGA) called *stop codons* (see Figure 6–50). These are not recognized by a tRNA and do not specify an amino acid, but instead signal

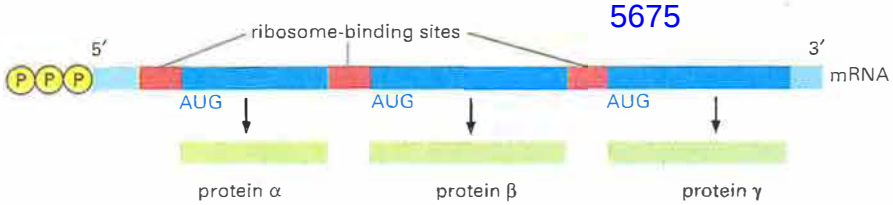


Figure 6–72 Structure of a typical bacterial mRNA molecule. Unlike eucaryotic ribosomes, which typically require a capped 5′ end, procaryotic ribosomes initiate transcription at ribosome-binding sites (Shine–Dalgarno sequences), which can be located anywhere along an mRNA molecule. This property of ribosomes permits bacteria to synthesize more than one type of protein from a single mRNA molecule.

to the ribosome to stop translation. Proteins known as *release factors* bind to any ribosome with a stop codon positioned in the A site, and this binding forces the peptidyl transferase in the ribosome to catalyze the addition of a water molecule instead of an amino acid to the peptidyl-tRNA (Figure 6–73). This reaction frees the carboxyl end of the growing polypeptide chain from its attachment to a tRNA molecule, and since only this attachment normally holds the growing polypeptide to the ribosome, the completed protein chain is immediately released into the cytoplasm. The ribosome then releases the mRNA and separates into the large and small subunits, which can assemble on another mRNA molecule to begin a new round of protein synthesis.

Release factors provide a dramatic example of *molecular mimicry*, whereby one type of macromolecule resembles the shape of a chemically unrelated molecule. In this case, the three-dimensional structure of release factors (made entirely of protein) bears an uncanny resemblance to the shape and charge distribution of a tRNA molecule (Figure 6–74). This shape and charge mimicry allows the release factor to enter the A-site on the ribosome and cause translation termination.

During translation, the nascent polypeptide moves through a large, water-filled tunnel (approximately 10 nm × 1.5 nm) in the large subunit of the ribosome (see Figure 6–68C). The walls of this tunnel, made primarily of 23S rRNA, are a patchwork of tiny hydrophobic surfaces embedded in a more extensive hydrophilic surface. This structure, because it is not complementary to any peptide structure, provides a "Teflon" coating through which a polypeptide chain can easily slide. The dimensions of the tunnel suggest that nascent proteins are largely unstructured as they pass through the ribosome, although some α-helical regions of the protein can form before leaving the ribosome tunnel. As it leaves the ribosome, a newly-synthesized protein must fold into its proper three-dimensional structure to be useful to the cell, and later in this chapter we discuss how this folding occurs. First, however, we review several additional aspects of the translation process itself.

Proteins Are Made on Polyribosomes

The synthesis of most protein molecules takes between 20 seconds and several minutes. But even during this very short period, multiple initiations usually take place on each mRNA molecule being translated. As soon as the preceding ribosome has translated enough of the nucleotide sequence to move out of the way, the 5′ end of the mRNA is threaded into a new ribosome. The mRNA molecules being translated are therefore usually found in the form of *polyribosomes* (also known as *polysomes*), large cytoplasmic assemblies made up of several ribosomes spaced as close as 80 nucleotides apart along a single mRNA molecule (Figure 6–75). These multiple initiations mean that many more protein

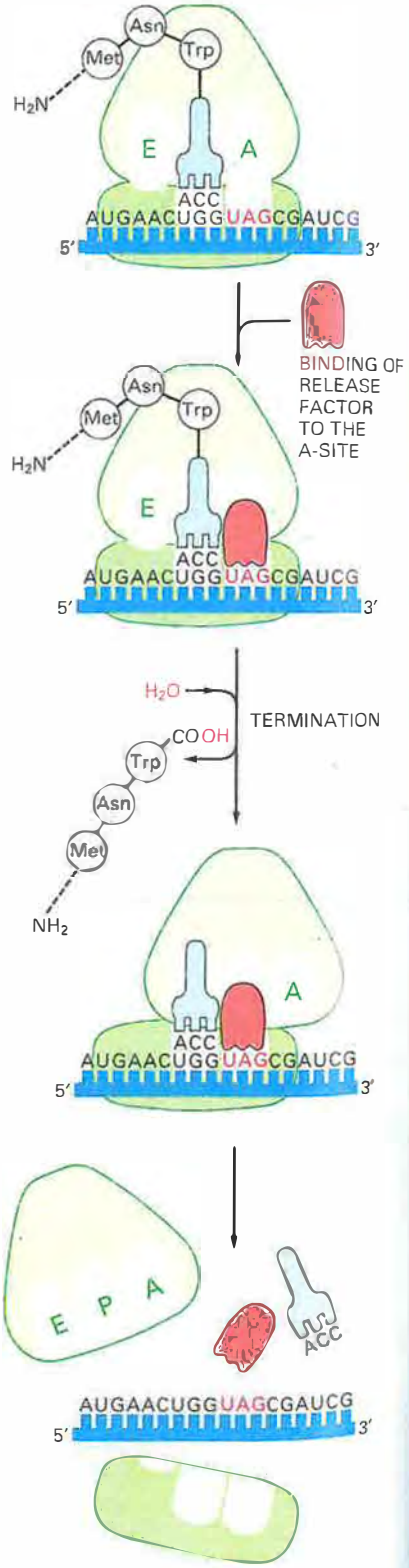


Figure 6–73 The final phase of protein synthesis. The binding of a release factor to an A-site bearing a stop codon terminates translation. The completed polypeptide is released and, after the action of a *ribosome recycling factor* (not shown), the ribosome dissociates into its two separate subunits.

Figure 6–74 The structure of a human translation release factor (eRF1) and its resemblance to a tRNA molecule. The protein is on the *left* and the tRNA on the *right*. (From H. Song et al., *Cell* 100:311–321, 2000. © Elsevier.)

molecules can be made in a given time than would be possible if each had to be completed before the next could start.

Both bacteria and eucaryotes utilize polysomes, and both employ additional strategies to speed up the rate of protein synthesis even further. Because bacterial mRNA does not need to be processed and is accessible to ribosomes while it is being made, ribosomes can attach to the free end of a bacterial mRNA molecule and start translating it even before the transcription of that RNA is complete, following closely behind the RNA polymerase as it moves along DNA. In eucaryotes, as we have seen, the 5′ and 3′ ends of the mRNA interact (see Figures 6–40 and 6–75A); therefore, as soon as a ribosome dissociates, its two subunits are in an optimal position to reinitiate translation on the same mRNA molecule.

Quality-Control Mechanisms Operate at Many Stages of Translation

Translation by the ribosome is a compromise between the opposing constraints of accuracy and speed. We have seen, for example, that the accuracy of translation (1 mistake per 10^4 amino acids joined) requires a time delay each time a new amino acid is added to a growing polypeptide chain, producing an overall

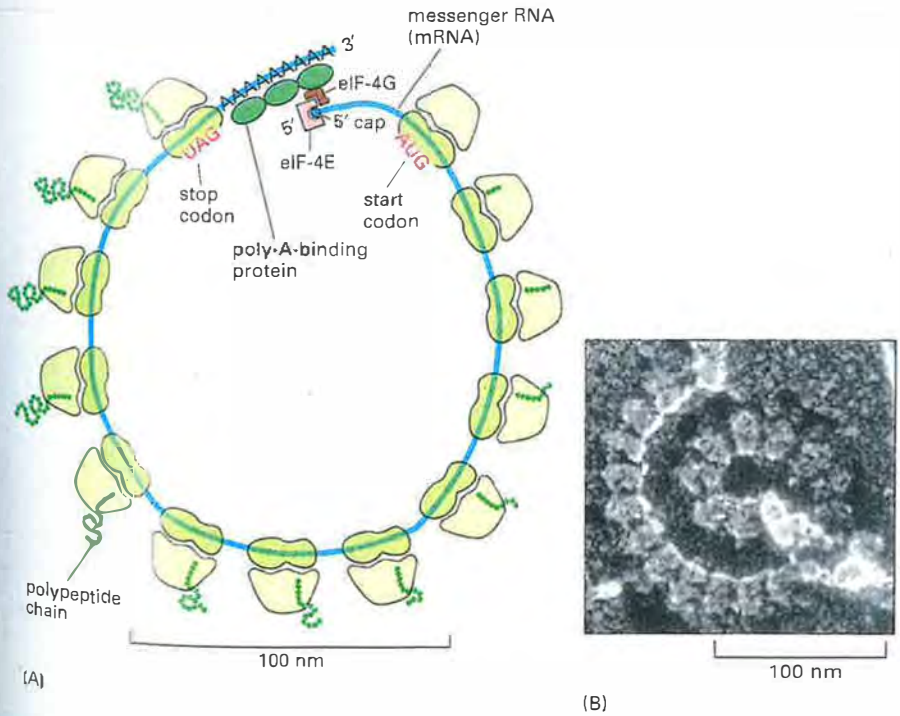


Figure 6–75 A polyribosome. (A) Schematic drawing showing how a series of ribosomes can simultaneously translate the same eucaryotic mRNA molecule. (B) Electron micrograph of a polyribosome from a eucaryotic cell. (B, courtesy of John Heuser.)

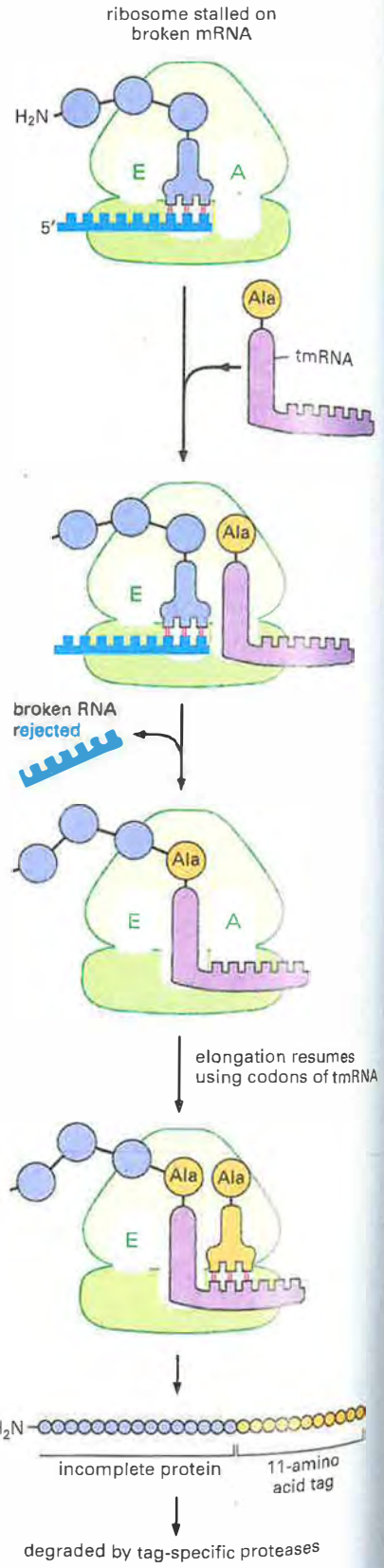


Figure 6–76 The rescue of a bacterial ribosome stalled on an incomplete mRNA molecule. The tmRNA shown is a 363-nucleotide RNA with both tRNA and mRNA functions, hence its name. It carries an alanine and can enter the vacant A-site of a stalled ribosome to add this alanine to a polypeptide chain, mimicking a tRNA except that no codon is present to guide it. The ribosome then translates ten codons from the tmRNA, completing an 11-amino acid tag on the protein. This tag is recognized by proteases that then degrade the entire protein.

speed of translation of 20 amino acids incorporated per second in bacteria. Mutant bacteria with a specific alteration in their small ribosomal subunit translate mRNA into protein with an accuracy considerably higher than this; however, protein synthesis is so slow in these mutants that the bacteria are barely able to survive.

We have also seen that attaining the observed accuracy of protein synthesis requires the expenditure of a great deal of free energy; this is expected, since, as discussed in Chapter 2, a price must be paid for any increase in order in the cell. In most cells, protein synthesis consumes more energy than any other biosynthetic process. At least four high-energy phosphate bonds are split to make each new peptide bond: two are consumed in charging a tRNA molecule with an amino acid (see Figure 6–56), and two more drive steps in the cycle of reactions occurring on the ribosome during synthesis itself (see Figure 6–66). In addition, extra energy is consumed each time that an incorrect amino acid linkage is hydrolyzed by a tRNA synthetase (see Figure 6–59) and each time that an incorrect tRNA enters the ribosome, triggers GTP hydrolysis, and is rejected (Figure 6–66). To be effective, these proofreading mechanisms must also remove an appreciable fraction of correct interactions; for this reason they are even more costly in energy than they might seem.

Other quality control mechanisms ensure that a eucaryotic mRNA molecule is complete before ribosomes even begin to translate it. Translating broken or partly processed mRNAs would be harmful to the cell, because truncated or otherwise aberrant proteins would be produced. In eucaryotes, we have seen that mRNA production involves not only transcription but also a series of elaborate RNA-processing steps; these take place in the nucleus, segregated from ribosomes, and only when the processing is complete are the mRNAs transported to the cytoplasm to be translated (see Figure 6–40). An mRNA molecule that was intact when it left the nucleus can, however, become broken in the cytosol. To avoid translating such broken mRNA molecules, the 5′ cap and the poly-A tail are both recognized by the translation-initiation apparatus before translation begins (see Figures 6–71 and 6–75).

Bacteria solve the problem of incomplete mRNAs in an entirely different way. Not only are there no signals at the 3′ ends of bacterial mRNAs, but also, as we have seen, translation often begins before the synthesis of the transcript has been completed. When the bacterial ribosome translates to the end of an incomplete RNA, a special RNA (called *tmRNA*) enters the A-site of the ribosome and is itself translated; this adds a special 11 amino acid tag to the C terminus of the truncated protein that signals to proteases that the entire protein is to be degraded (Figure 6–76).

There Are Minor Variations in the Standard Genetic Code

As discussed in Chapter 1, the genetic code (shown in Figure 6–50) applies to all three major branches of life, providing important evidence for the common ancestry of all life on Earth. Although rare, there are exceptions to this code, and we discuss some of them in this section. For example, *Candida albicans*, the most prevalent human fungal pathogen, translates the codon CUG as serine, whereas nearly all other organisms translate it as leucine. Mitochondria (which have their own genomes and encode much of their translational apparatus) also show several deviations from the standard code. For example, in mammalian mitochondria AUA is translated as methionine, whereas in the cytosol of the cell it is translated as isoleucine (see Table 14–3, p. 814).

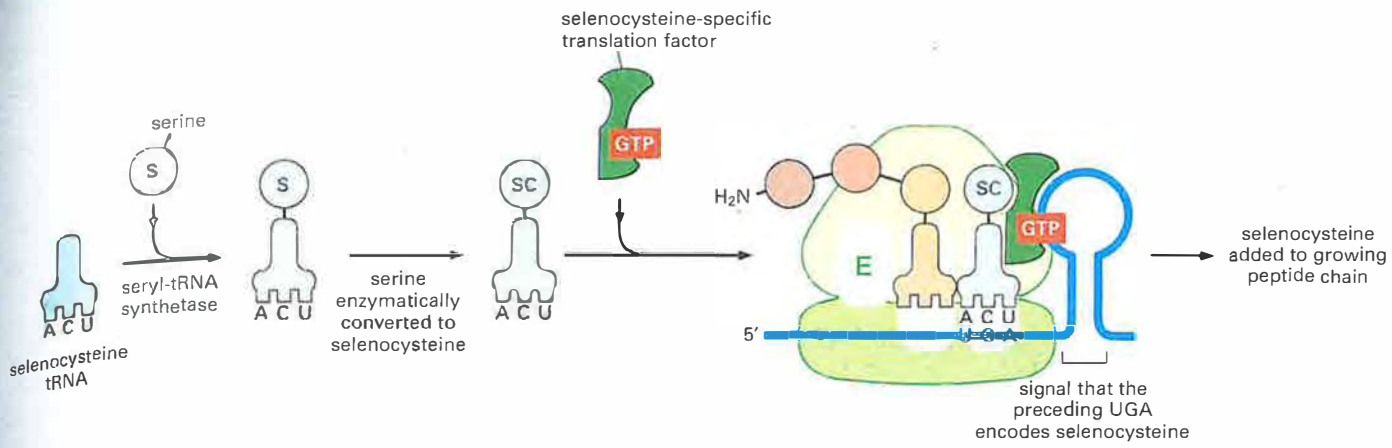


Figure 6–77 Incorporation of selenocysteine into a growing polypeptide chain. A specialized tRNA is charged with serine by the normal seryl-tRNA synthetase, and the serine is subsequently converted enzymatically to selenocysteine. A specific RNA structure in the mRNA (a stem and loop structure with a particular nucleotide sequence) signals that selenocysteine is to be inserted at the neighboring UGA codon. As indicated, this event requires the participation of a selenocysteine-specific translation factor.

The type of deviation in the genetic code discussed above is "hardwired" into the organisms or the organelles in which it occurs. A different type of variation, sometimes called *translational recoding*, occurs in many cells. In this case, other nucleotide sequence information present in an mRNA can change the meaning of the genetic code at a particular site in the mRNA molecule. The standard code allows cells to manufacture proteins using only 20 amino acids. However, bacteria, archaea, and eucaryotes have available to them a twenty-first amino acid that can be incorporated directly into a growing polypeptide chain through translational recoding. Selenocysteine, which is essential for the efficient function of a variety of enzymes, contains a selenium atom in place of the sulfur atom of cysteine. Selenocysteine is produced from a serine attached to a special tRNA molecule that base-pairs with the UGA codon, a codon normally used to signal a translation stop. The mRNAs for proteins in which selenocysteine is to be inserted at a UGA codon carry an additional nucleotide sequence in the mRNA nearby that causes this recoding event (Figure 6–77).

Another form of recoding is *translational frameshifting*. This type of recoding is commonly used by retroviruses, a large group of eucaryotic viruses, in which it allows more than one protein to be synthesized from a single mRNA. These viruses commonly make both the capsid proteins *(Gag proteins)* and the viral reverse transcriptase and integrase *(Pol proteins)* from the same RNA transcript (see Figure 5–73). Such a virus needs many more copies of the Gag proteins than it does of the Pol proteins, and they achieve this quantitative adjustment by encoding the *pol* genes just after the *gag* genes but in a different reading frame. A stop codon at the end of the *gag* coding sequence can be bypassed on occasion by an intentional translational frameshift that occurs upstream of it. This frameshift occurs at a particular codon in the mRNA and requires a specific *recoding signal*, which seems to be a structural feature of the RNA sequence downstream of this site (Figure 6–78).

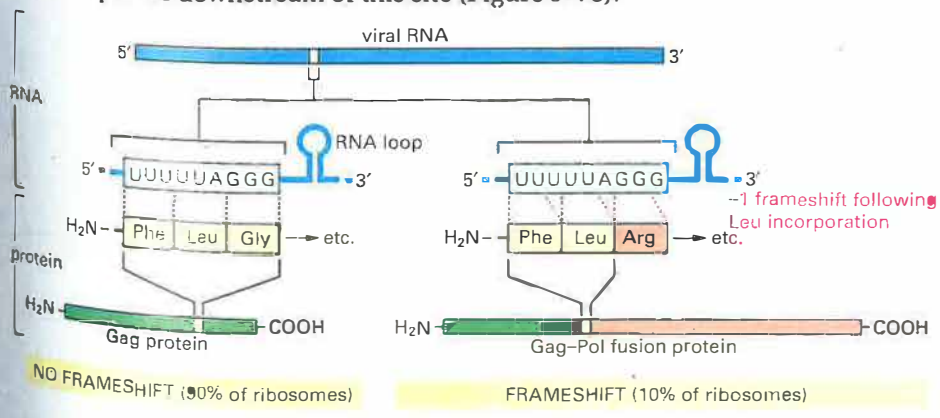


Figure 6–78 The translational frameshifting that produces the reverse transcriptase and integrase of a retrovirus. The viral reverse transcriptase and integrase are produced by proteolytic processing of a large protein (the Gag–Pol fusion protein) consisting of both the Gag and Pol amino acid sequences. The viral capsid proteins are produced by proteolytic processing of the more abundant Gag protein. Both the Gag and the Gag–Pol fusion proteins start identically, but the Gag protein terminates at an in-frame stop codon (not shown); the indicated frameshift bypasses this stop codon, allowing the synthesis of the longer Gag–Pol fusion protein. The frameshift occurs because features in the local RNA structure (including the RNA loop shown) cause the $tRNA^{Leu}$ attached to the C-terminus of the growing polypeptide chain occasionally to slip backward by one nucleotide on the ribosome, so that it pairs with a UUU codon instead of the UUA codon that had initially specified its incorporation; the next codon (AGG) in the new reading frame specifies an arginine rather than a glycine. This controlled slippage is due in part to a stem and loop structure that forms in the viral mRNA, as indicated in the figure. The sequence shown is from the human AIDS virus, HIV. (Adapted from T. Jacks et al., *Nature* 331:280–283, 1988.)

Many Inhibitors of Procaryotic Protein Synthesis Are Useful as Antibiotics

Many of the most effective antibiotics used in modern medicine are compounds made by fungi that act by inhibiting bacterial protein synthesis. Some of these drugs exploit the structural and functional differences between bacterial and eucaryotic ribosomes so as to interfere preferentially with the function of bacterial ribosomes. Thus some of these compounds can be taken in high doses without undue toxicity to humans. Because different antibiotics bind to different regions of bacterial ribosomes, they often inhibit different steps in the synthetic process. Some of the more common antibiotics of this kind are listed in Table 6–3 along with several other inhibitors of protein synthesis, some of which act on eucaryotic cells and therefore cannot be used as antibiotics.

Because they block specific steps in the processes that lead from DNA to protein, many of the compounds listed in Table 6–3 are useful for cell biological studies. Among the most commonly used drugs in such experimental studies are *chloramphenicol*, *cycloheximide*, and *puromycin*, all of which specifically inhibit protein synthesis. In a eucaryotic cell, for example, chloramphenicol inhibits protein synthesis on ribosomes only in mitochondria (and in chloroplasts in plants), presumably reflecting the procaryotic origins of these organelles (discussed in Chapter 14). Cycloheximide, in contrast, affects only ribosomes in the cytosol. Puromycin is especially interesting because it is a structural analog of a tRNA molecule linked to an amino acid and is therefore another example of molecular mimicry; the ribosome mistakes it for an authentic amino acid and covalently incorporates it at the C-terminus of the growing peptide chain, thereby causing the premature termination and release of the polypeptide. As might be expected, puromycin inhibits protein synthesis in both procaryotes and eucaryotes.

Having described the translation process itself, we now discuss how its products—the proteins of the cell—fold into their correct three-dimensional conformations.

A Protein Begins to Fold While It Is Still Being Synthesized

The process of gene expression is not over when the genetic code has been used to create the sequence of amino acids that constitutes a protein. To be useful to the cell, this new polypeptide chain must fold up into its unique three-dimensional

TABLE 6–3 Inhibitors of Protein or RNA Synthesis

INHIBITOR	SPECIFIC EFFECT
Acting only on bacteria	
Tetracycline	blocks binding of aminoacyl-tRNA to A-site of ribosome
Streptomycin	prevents the transition from initiation complex to chain-elongating ribosome and also causes miscoding
Chloramphenicol	blocks the peptidyl transferase reaction on ribosomes (step 2 in Figure 6–65)
Erythromycin	blocks the translocation reaction on ribosomes (step 3 in Figure 6–65)
Rifamycin	blocks initiation of RNA chains by binding to RNA polymerase (prevents RNA synthesis)
Acting on bacteria and eucaryotes	
Puromycin	causes the premature release of nascent polypeptide chains by its addition to growing chain end
Actinomycin D	binds to DNA and blocks the movement of RNA polymerase (prevents RNA synthesis)
Acting on eucaryotes but not bacteria	
Cycloheximide	blocks the translocation reaction on ribosomes (step 3 in Figure 6–65)
Anisomycin	blocks the peptidyl transferase reaction on ribosomes (step 2 in Figure 6–65)
α-Amanitin	blocks mRNA synthesis by binding preferentially to RNA polymerase II

The ribosomes of eucaryotic mitochondria (and chloroplasts) often resemble those of bacteria in their sensitivity to inhibitors. Therefore, some of these antibiotics can have a deleterious effect on human mitochondria.

conformation, bind any small-molecule cofactors required for its activity, be appropriately modified by protein kinases or other protein-modifying enzymes, and assemble correctly with the other protein subunits with which it functions (Figure 6–79).

The information needed for all of the protein maturation steps listed above is ultimately contained in the sequence of linked amino acids that the ribosome produces when it translates an mRNA molecule into a polypeptide chain. As discussed in Chapter 3, when a protein folds into a compact structure, it buries most of its hydrophobic residues in an interior core. In addition, large numbers of noncovalent interactions form between various parts of the molecule. It is the sum of all of these energetically favorable arrangements that determines the final folding pattern of the polypeptide chain—as the conformation of lowest free energy (see p. 134).

Through many millions of years of evolutionary time, the amino acid sequence of each protein has been selected not only for the conformation that it adopts but also for an ability to fold rapidly, as its polypeptide chain spins out of the ribosome starting from the N-terminal end. Experiments have demonstrated that once a protein domain in a multi-domain protein emerges from the ribosome, it forms a compact structure within a few seconds that contains most of the final secondary structure (α helices and β sheets) aligned in roughly the right way (Figure 6–80). For many protein domains, this unusually open and flexible structure, which is called a *molten globule*, is the starting point for a relatively slow process in which many side-chain adjustments occur that eventually form the correct tertiary structure. Nevertheless, because it takes several minutes to synthesize a protein of average size, a great deal of the folding process is complete by the time the ribosome releases the C-terminal end of a protein (Figure 6–81).

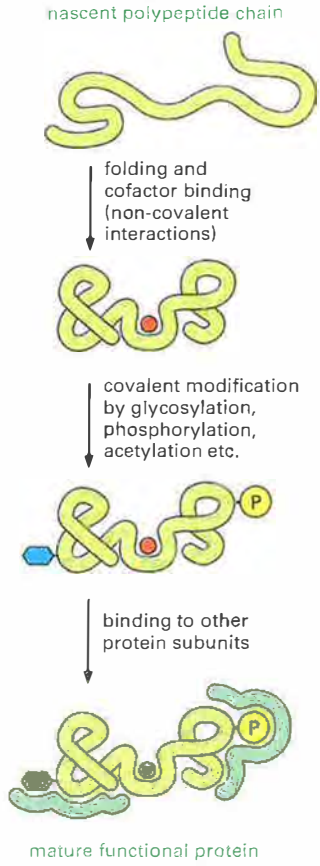


Figure 6–79 Steps in the creation of a functional protein. As indicated, translation of an mRNA sequence into an amino acid sequence on the ribosome is not the end of the process of forming a protein. To be useful to the cell, the completed polypeptide chain must fold correctly into its three-dimensional conformation, bind any cofactors required, and assemble with its partner protein chains (if any). These changes are driven by noncovalent bond formation. As indicated, many proteins also have covalent modifications made to selected amino acids. Although the most frequent of these are protein glycosylation and protein phosphorylation, more than 100 different types of covalent modifications are known (see, for example, Figure 4–35).

Molecular Chaperones Help Guide the Folding of Many Proteins

The folding of many proteins is made more efficient by a special class of proteins called **molecular chaperones**. The latter proteins are useful for cells because there are a variety of different paths that can be taken to convert the molten globule form of a protein to the protein's final compact conformation. For many proteins, some of the intermediates formed along the way would aggregate and be left as off-pathway dead ends without the intervention of a chaperone that resets the folding process (Figure 6–82).

Molecular chaperones were first identified in bacteria when *E. coli* mutants that failed to allow bacteriophage lambda to replicate in them were studied.

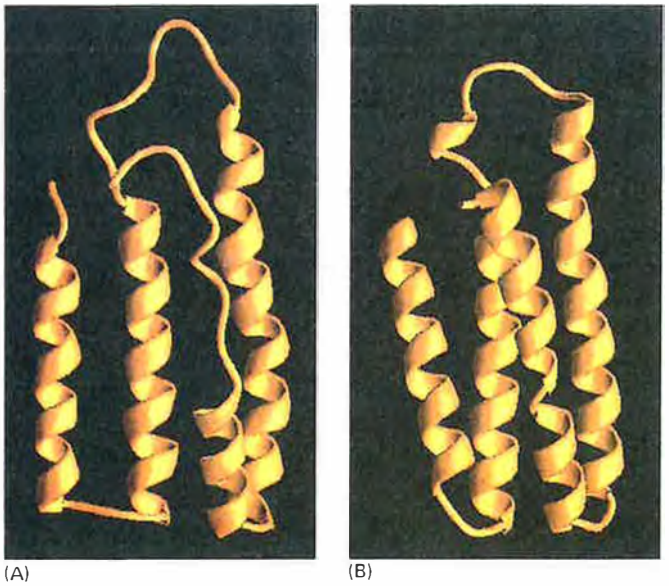

(A) (B)

Figure 6–80 The structure of a molten globule. (A) A molten globule form of cytochrome b_{562} is more open and less highly ordered than the final folded form of the protein, shown in (B). Note that the molten globule contains most of the secondary structure of the final form, although the ends of the α helices are frayed and one of the helices is only partly formed. (Courtesy of Joshua Wand, from Y. Feng et al., *Nat. Struct. Biol.* 1:30–35, 1994.)

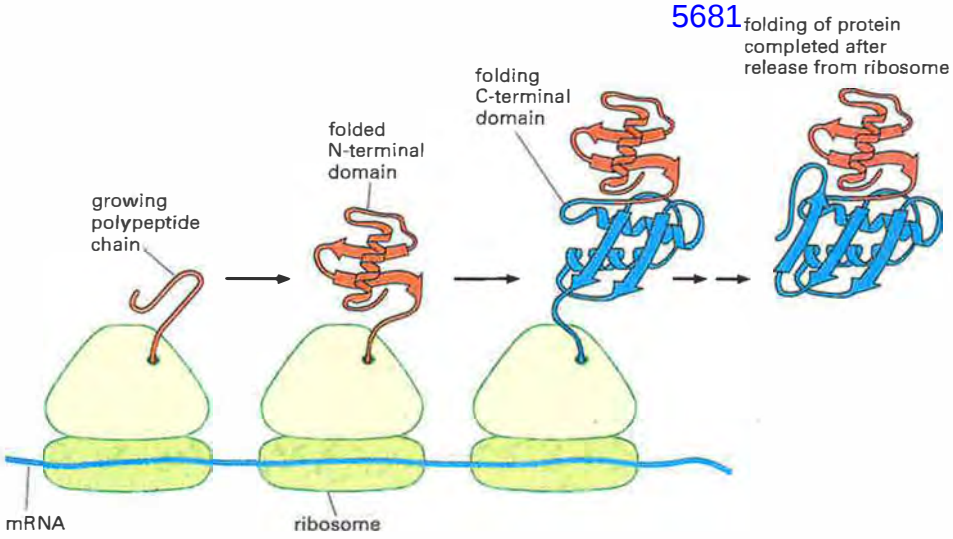


Figure 6–81 The co-translational folding of a protein. A growing polypeptide chain is shown acquiring its secondary and tertiary structure as it emerges from a ribosome. The N-terminal domain folds first, while the C-terminal domain is still being synthesized. In this case, the protein has not yet achieved its final conformation by the time it is released from the ribosome. (Modified from A.N. Federov and T.O. Baldwin, *J. Biol. Chem.* 272:32715–32718, 1997.)

These mutant cells produce slightly altered versions of the chaperone machinery, and as a result they are defective in specific steps in the assembly of the viral proteins. The molecular chaperones are included among the *heat-shock proteins* (hence their designation as *hsp*), because they are synthesized in dramatically increased amounts after a brief exposure of cells to an elevated temperature (for example, 42°C for cells that normally live at 37°C). This reflects the operation of a feedback system that responds to any increase in misfolded proteins (such as those produced by elevated temperatures) by boosting the synthesis of the chaperones that help these proteins refold.

Eucaryotic cells have at least two major families of molecular chaperones—known as the hsp60 and hsp70 proteins. Different family members function in different organelles. Thus, as discussed in Chapter 12, mitochondria contain their own hsp60 and hsp70 molecules that are distinct from those that function in the cytosol, and a special hsp70 (called *BIP*) helps to fold proteins in the endoplasmic reticulum.

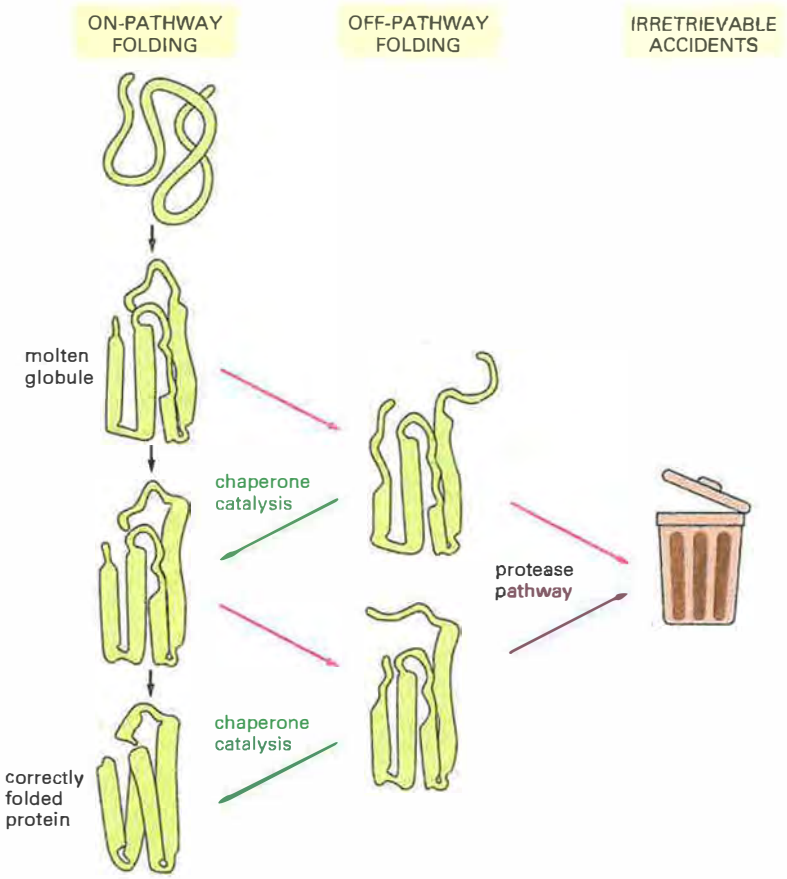


Figure 6–82 A current view of protein folding. Each domain of a newly synthesized protein rapidly attains a "molten globule" state. Subsequent folding occurs more slowly and by multiple pathways, often involving the help of a molecular chaperone. Some molecules may still fail to fold correctly; as explained shortly, these are recognized and degraded by specific proteases.

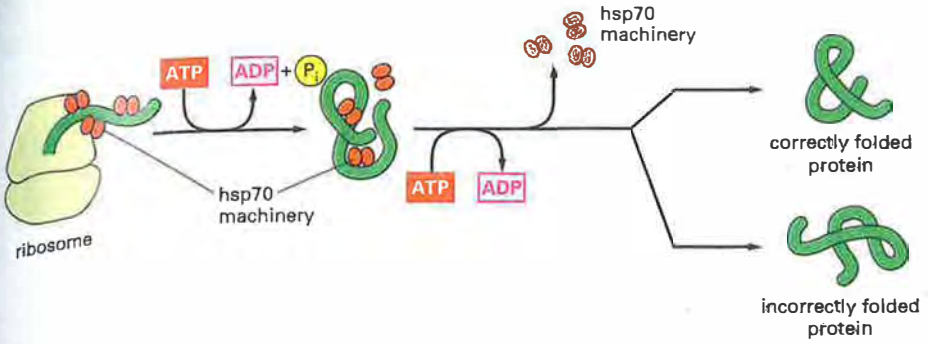


Figure 6–83 The hsp70 family of molecular chaperones. These proteins act early, recognizing a small stretch of hydrophobic amino acids on a protein's surface. Aided by a set of smaller hsp40 proteins, an hsp70 monomer binds to its target protein and then hydrolyzes a molecule of ATP to ADP, undergoing a conformational change that causes the hsp70 to clamp down very tightly on the target. After the hsp40 dissociates, the dissociation of the hsp70 protein is induced by the rapid re-binding of ATP after ADP release. Repeated cycles of hsp protein binding and release help the target protein to refold, as schematically illustrated in Figure 6–82.

The hsp60-like and hsp70 proteins each work with their own small set of associated proteins when they help other proteins to fold. They share an affinity for the exposed hydrophobic patches on incompletely folded proteins, and they hydrolyze ATP, often binding and releasing their protein with each cycle of ATP hydrolysis. In other respects, the two types of hsp proteins function differently. The hsp70 machinery acts early in the life of many proteins, binding to a string of about seven hydrophobic amino acids before the protein leaves the ribosome (Figure 6–83). In contrast, hsp60-like proteins form a large barrel-shaped structure that acts later in a protein's life, after it has been fully synthesized. This type of chaperone forms an "isolation chamber" into which misfolded proteins are fed, preventing their aggregation and providing them with a favorable environment in which to attempt to refold (Figure 6–84).

Exposed Hydrophobic Regions Provide Critical Signals for Protein Quality Control

If radioactive amino acids are added to cells for a brief period, the newly synthesized proteins can be followed as they mature into their final functional form. It is this type of experiment that shows that the hsp70 proteins act first, beginning when a protein is still being synthesized on a ribosome, and that the

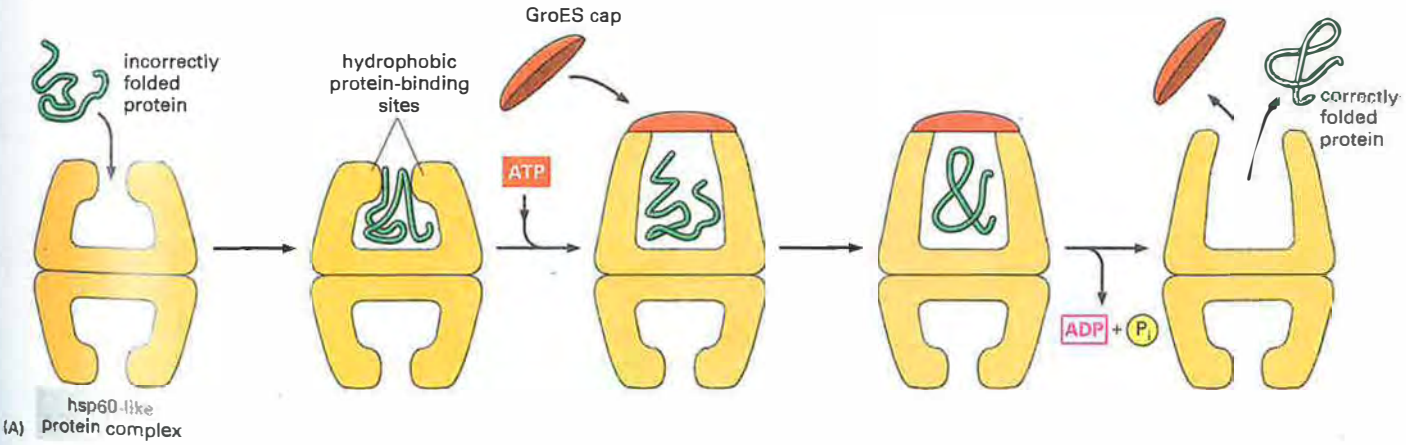
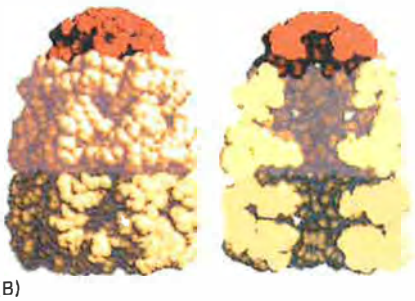


Figure 6–84 The structure and function of the hsp60 family of molecular chaperones. (A) The catalysis of protein refolding. As indicated, a misfolded protein is initially captured by hydrophobic interactions along one rim of the barrel. The subsequent binding of ATP plus a protein cap increases the diameter of the barrel rim, which may transiently stretch (partly unfold) the client protein. This also confines the protein in an enclosed space, where it has a new opportunity to fold. After about 15 seconds, ATP hydrolysis ejects the protein, whether folded or not, and the cycle repeats. This type of molecular chaperone is also known as a chaperonin; it is designated as hsp60 in mitochondria, TCP-1 in the cytosol of vertebrate cells, and GroEL in bacteria. As indicated, only half of the symmetrical barrel operates on a client protein at any one time. (B) The structure of GroEL bound to its GroES cap, as determined by x-ray crystallography. On the *left* is shown the outside of the barrel-like structure and on the *right* a cross section through its center. (B, adapted from B. Bukace and A.L. Horwich, *Cell* 92:351–366, 1998.)


(B)

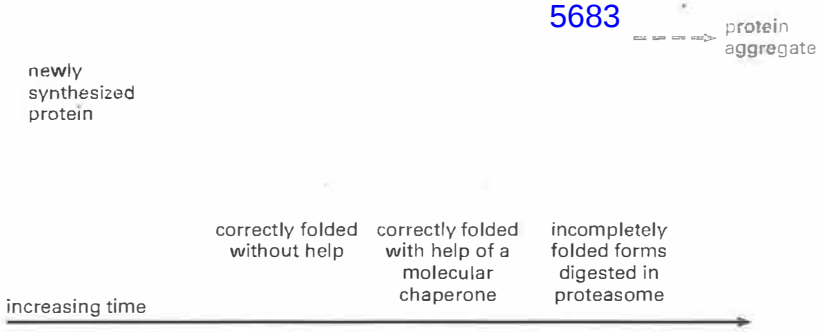


Figure 6–85 The cellular mechanisms that monitor protein quality after protein synthesis. As indicated, a newly synthesized protein sometimes folds correctly and assembles with its partners on its own, in which case it is left alone. Incompletely folded proteins are helped to refold by molecular chaparones: first by a family of hsp70 proteins, and if this fails, then by hsp60-like proteins. In both cases the client proteins are recognized by an abnormally exposed patch of hydrophobic amino acids on their surface. These processes compete with a different system that recognizes an abnormally exposed patch and transfers the protein that contains it to a proteasome for complete destruction. The combination of all of these processes is needed to prevent massive protein aggregation in a cell, which can occur when many hydrophobic regions on proteins clump together and precipitate the entire mass out of solution.

hsp60-like proteins are called into play only later to help in folding completed proteins. However, the same experiments reveal that only a subset of the newly synthesized proteins becomes involved: perhaps 20% of all proteins with the hsp70 and 10% with the hsp60-like molecular chaperones. How are these proteins selected for this ATP-catalyzed refolding?

Before answering, we need to pause to consider the post-translational fate of proteins more broadly. A protein that has a sizable exposed patch of hydrophobic amino acids on its surface is usually abnormal: it has either failed to fold correctly after leaving the ribosome, suffered an accident that partly unfolded it at a later time, or failed to find its normal partner subunit in a larger protein complex. Such a protein is not merely useless to the cell, it can be dangerous. Many proteins with an abnormally exposed hydrophobic region can form large aggregates, precipitating out of solution. We shall see that, in rare cases, such aggregates do form and cause severe human diseases. But in the vast majority of cells, powerful protein quality control mechanisms prevent such disasters.

Given this background, it is not surprising that cells have evolved elaborate mechanisms that recognize and remove the hydrophobic patches on proteins. Two of these mechanisms depend on the molecular chaperones just discussed, which bind to the patch and attempt to repair the defective protein by giving it another chance to fold. At the same time, by covering the hydrophobic patches, these chaperones transiently prevent protein aggregation. Proteins that very rapidly fold correctly on their own do not display such patches and are therefore bypassed by chaperones.

Figure 6–85 outlines all of the quality control choices that a cell makes for a difficult-to-fold, newly synthesized protein. As indicated, when attempts to refold a protein fail, a third mechanism is called into play that completely destroys the protein by proteolysis. The proteolytic pathway begins with the recognition of an abnormal hydrophobic patch on a protein's surface, and it ends with the delivery of the entire protein to a protein destruction machine, a complex protease known as the *proteasome*. As described next, this process depends on an elaborate protein-marking system that also carries out other central functions in the cell by destroying selected normal proteins.

The Proteasome Degrades a Substantial Fraction of the Newly Synthesized Proteins in Cells

Cells quickly remove the failures of their translation processes. Recent experiments suggest that as many as one-third of the newly made polypeptide chains are selected for rapid degradation as a result of the protein quality control mechanisms just described. The final disposal apparatus in eucaryotes is the **proteasome**, an abundant ATP-dependent protease that constitutes nearly 1% of cellular protein. Present in many copies dispersed throughout the cytosol and the nucleus, the proteasome also targets proteins of the endoplasmic reticulum (ER): those proteins that fail either to fold or to be assembled properly after they enter the ER are detected by an ER-based surveillance system that *retrotranslocate* them back to the cytosol for degradation (discussed in Chapter 12).

Each proteasome consists of a central hollow cylinder (the 20S core proteasome) formed from multiple protein subunits that assemble as a cylindrical

stack of four heptameric rings. Some of these subunits are distinct proteases whose active sites face the cylinder's inner chamber (Figure 6–86A). Each end of the cylinder is normally associated with a large protein complex (the 19S cap) containing approximately 20 distinct polypeptides (Figure 6–86B). The cap subunits include at least six proteins that hydrolyze ATP; located near the edge of the cylinder, these ATPases are thought to unfold the proteins to be digested and move them into the interior chamber for proteolysis. A crucial property of the proteasome, and one reason for the complexity of its design, is the *processivity* of its mechanism: in contrast to a "simple" protease that cleaves a substrate's polypeptide chain just once before dissociating, the proteasome keeps the entire substrate bound until all of it is converted into short peptides.

The 19S caps act as regulated "gates" at the entrances to the inner proteolytic chamber, being also responsible for binding a targeted protein substrate to the proteasome. With a few exceptions, the proteasomes act on proteins that have been specifically marked for destruction by the covalent attachment of multiple copies of a small protein called *ubiquitin* (Figure 6–87A). Ubiquitin exists in cells either free or covalently linked to a huge variety of intracellular proteins. For most of these proteins, this tagging by ubiquitin results in their destruction by the proteasome.

An Elaborate Ubiquitin-conjugating System Marks Proteins for Destruction

Ubiquitin is prepared for conjugation to other proteins by the ATP-dependent *ubiquitin-activating* enzyme (E1), which creates an activated ubiquitin that is transferred to one of a set of ubiquitin-conjugating (E2) enzymes. The E2 enzymes act in conjunction with accessory (E3) proteins. In the E2–E3 complex, called *ubiquitin ligase*, the E3 component binds to specific degradation signals in protein substrates, helping E2 to form a *multiubiquitin* chain linked to a lysine of the substrate protein. In this chain, the C-terminal residue of each ubiquitin is linked to a specific lysine of the preceding ubiquitin molecule, producing a linear series of ubiquitin–ubiquitin conjugates (Figure 6–87B). It is this multiubiquitin chain on a target protein that is recognized by a specific receptor in the proteasome.

There are roughly 30 structurally similar but distinct E2 enzymes in mammals, and hundreds of different E3 proteins that form complexes with specific E2 enzymes. The ubiquitin–proteasome system thus consists of many distinct but similarly organized proteolytic pathways, which have in common both the E1 enzyme at the "top" and the proteasome at the "bottom," and differ by the compositions of their E2–E3 ubiquitin ligases and accessory factors. Distinct ubiquitin ligases recognize different degradation signals, and therefore target for degradation distinct subsets of intracellular proteins that bear these signals.

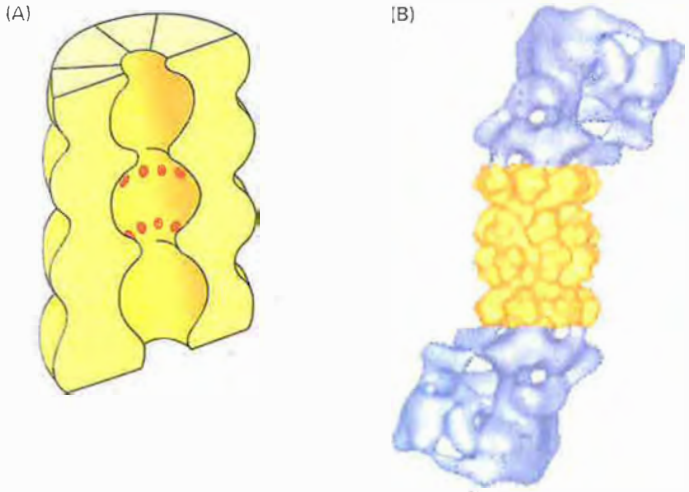
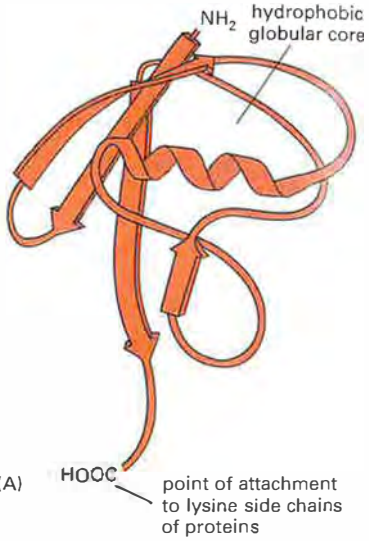


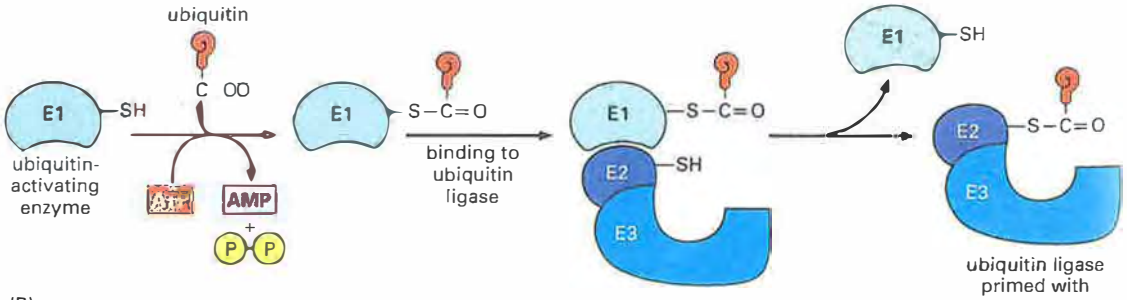
Figure 6–86 The proteasome. (A) A cut-away view of the structure of the central 20S cylinder, as determined by x-ray crystallography, with the active sites of the proteases indicated by *red dots*. (B) The structure of the entire proteasome, in which the central cylinder *(yellow)* is supplemented by a 19S cap *(blue)* at each end, whose structure has been determined by computer processing of electron microscope images. The complex cap structure selectively binds those proteins that have been marked for destruction; it then uses ATP hydrolysis to unfold their polypeptide chains and feed them into the inner chamber of the 20S cylinder for digestion to short peptides. (B, from W. Baumeister et al., *Cell* 92:367–380, 1998. © Elsevier.)

Denatured or otherwise misfolded proteins, as well as proteins containing oxidized or other abnormal amino acids, are recognized and destroyed because abnormal proteins tend to present on their surface amino acid sequences or conformational motifs that are recognized as degradation signals by a set of E3 molecules in the ubiquitin–proteasome system; these sequences must of course be buried and therefore inaccessible in the normal counterparts of these proteins. However, a proteolytic pathway that recognizes and destroys abnormal proteins must be able to distinguish between *completed* proteins that have "wrong" conformations and the many growing polypeptides on ribosomes (as well as polypeptides just released from ribosomes) that have not yet achieved their normal folded conformation. This is not a trivial problem; the ubiquitin–proteasome system is thought to destroy some of the nascent and newly formed protein molecules not because these proteins are abnormal as such but because they transiently expose degradation signals that are buried in their mature (folded) state.

Many Proteins Are Controlled by Regulated Destruction

One function of intracellular proteolytic mechanisms is to recognize and eliminate misfolded or otherwise abnormal proteins, as just described. Yet another function of these proteolytic pathways is to confer short half-lives on specific normal proteins whose concentrations must change promptly with alterations





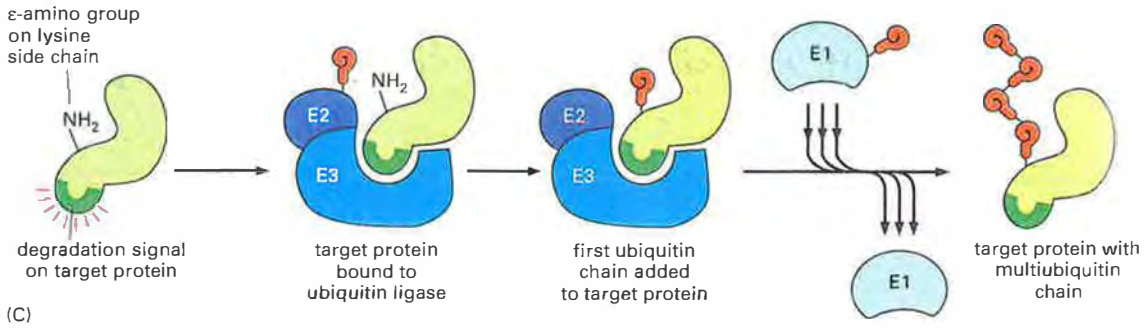


Figure 6–87 Ubiquitin and the marking of proteins with multiubiquitin chains. (A) The three-dimensional structure of ubiquitin; this relatively small protein contains 76 amino acids. (B) The C-terminus of ubiquitin is initially activated through its high-energy thioester linkage to a cysteine side chain on the E1 protein. This reaction requires ATP, and it proceeds via a covalent AMP-ubiquitin intermediate. The activated ubiquitin on E1, also known as the ubiquitin-activating enzyme, is then transferred to the cysteines on a set of E2 molecules. These E2s exist as complexes with an even larger family of E3 molecules. (C) The addition of a multiubiquitin chain to a target protein. In a mammalian cell there are roughly 300 distinct E2–E3 complexes, each of which recognizes a different degradation signal on a target protein by means of its E3 component. The E2s are called ubiquitin-conjugating enzymes. The E3s have been referred to traditionally as ubiquitin ligases, but it is more accurate to reserve this name for the functional E2–E3 complex.

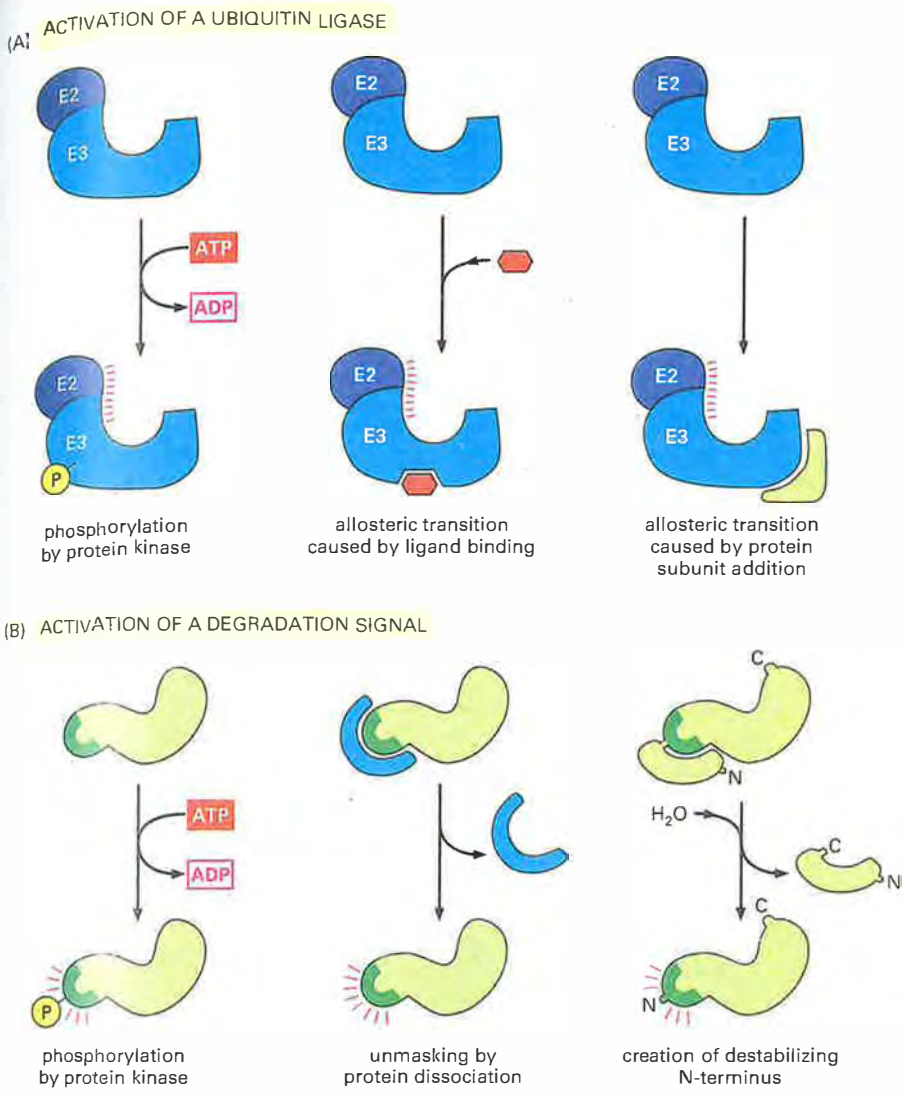


Figure 6–88 Two general ways of inducing the degradation of a specific protein. (A) Activation of a specific E3 molecule creates a new ubiquitin ligase. (B) Creation of an exposed degradation signal in the protein to be degraded. This signal binds a ubiquitin ligase, causing the addition of a multiubiquitin chain to a nearby lysine on the target protein. All six pathways shown are known to be used by cells to induce the movement of selected proteins into the proteasome.

in the state of a cell. Some of these short-lived proteins are degraded rapidly at all times, while many others are *conditionally* short-lived, that is, they are metabolically stable under some conditions, but become unstable upon a change in the cell's state. For example, mitotic cyclins are long-lived throughout the cell cycle until their sudden degradation at the end of mitosis, as explained in Chapter 17.

How is such a regulated destruction of a protein controlled? A variety of mechanisms are known, as illustrated through specific examples later in this book. In one general class of mechanism (Figure 6–88A), the activity of a ubiquitin ligase is turned on either by E3 phosphorylation or by an allosteric transition in an E3 protein caused by its binding to a specific small or large molecule. For example, the anaphase-promoting complex (APC) is a multisubunit ubiquitin ligase that is activated by a cell-cycle-timed subunit addition at mitosis. The activated APC then causes the degradation of mitotic cyclins and several other regulators of the metaphase–anaphase transition (see Figure 17–20).

Alternatively, in response either to intracellular signals or to signals from the environment, a degradation signal can be created in a protein, causing its rapid ubiquitylation and destruction by the proteasome. One common way to create such a signal is to phosphorylate a specific site on a protein that unmasks a normally hidden degradation signal. Another way to unmask such a signal is by the regulated dissociation of a protein subunit. Finally, powerful degradation signals can be created by a single cleavage of a peptide bond, provided that this cleavage creates a new N-terminus that is recognized by a specific E3 as a "destabilizing" N-terminal residue (Figure 6–88B).

The N-terminal type of degradation signal arises because of the "N-end rule," which relates the half-life of a protein *in vivo* to the identity of its N-terminal residue. There are 12 destabilizing residues in the N-end rule of the yeast *S. cerevisiae* (Arg, Lys, His, Phe, Leu, Tyr, Trp, Ile, Asp, Glu, Asn, and Gln), out of the 20 standard amino acids. The destabilizing N-terminal residues are recognized by a special ubiquitin ligase that is conserved from yeast to humans.

As we have seen, all proteins are initially synthesized bearing methionine (or formylmethionine in bacteria), as their N-terminal residue, which is a stabilizing residue in the N-end rule. Special proteases, called methionine aminopeptidases, will often remove the first methionine of a nascent protein, but they will do so only if the second residue is also stabilizing in the yeast-type N-end rule. Therefore, it was initially unclear how N-end rule substrates form *in vivo*. However, it has recently been discovered that a subunit of cohesin, a protein complex that holds sister chromatids together, is cleaved by a site-specific protease at the metaphase–anaphase transition. This cell-cycle-regulated cleavage allows separation of the sister chromatids and leads to the completion of mitosis (see Figure 17–26). The C-terminal fragment of the cleaved subunit bears an N-terminal arginine, a destabilizing residue in the N-end rule. Mutant cells lacking the N-end rule pathway exhibit a greatly increased frequency of chromosome loss, presumably because a failure to degrade this fragment of the cohesin subunit interferes with the formation of new chromatid-associated cohesin complexes in the next cell cycle.

Abnormally Folded Proteins Can Aggregate to Cause Destructive Human Diseases

When all of a cell's protein quality controls fail, large protein aggregates tend to accumulate in the affected cell (see Figure 6–85). Some of these aggregates, by adsorbing critical macromolecules to them, can severely damage cells and even cause cell death. The protein aggregates released from dead cells tend to accumulate in the extracellular matrix that surrounds the cells in a tissue, and in extreme cases they can also damage tissues. Because the brain is composed of a highly organized collection of nerve cells, it is especially vulnerable. Not surprisingly, therefore, protein aggregates primarily cause diseases of neurodegeneration. Prominent among these are Huntington's disease and Alzheimer's disease—the latter causing age-related dementia in more than 20 million people in today's world.

For a particular type of protein aggregate to survive, grow, and damage an organism, it must be highly resistant to proteolysis both inside and outside the cell. Many of the protein aggregates that cause problems form fibrils built from a series of polypeptide chains that are layered one over the other as a continuous stack of β sheets. This so-called *cross-beta filament* (Figure 6–89C) tends to be highly resistant to proteolysis. This resistance presumably explains why this structure is observed in so many of the neurological disorders caused by protein aggregates, where it produces abnormally staining deposits known as *amyloid*.

One particular variety of these diseases has attained special notoriety. These are the **prion diseases**. Unlike Huntington's or Alzheimer's disease, a prion disease can spread from one organism to another, providing that the second organism eats a tissue containing the protein aggregate. A set of diseases—called scrapie in sheep, Creutzfeldt–Jacob disease (CJD) in humans, and bovine spongiform encephalopathy (BSE) in cattle—are caused by a misfolded, aggregated form of a protein called PrP (for prion protein). The PrP is normally located on the outer surface of the plasma membrane, most prominently in neurons. Its normal function is not known. However, PrP has the unfortunate property of being convertible to a very special abnormal conformation (Figure 6–89A). This conformation not only forms protease-resistant, cross-beta filaments; it also is "infectious" because it converts normally folded molecules of PrP to the same form. This property creates a positive feedback loop that propagates the abnormal form of PrP, called PrP* (Figure 6–89B) and thereby allows PrP to spread rapidly from cell to cell in the brain, causing the death of both

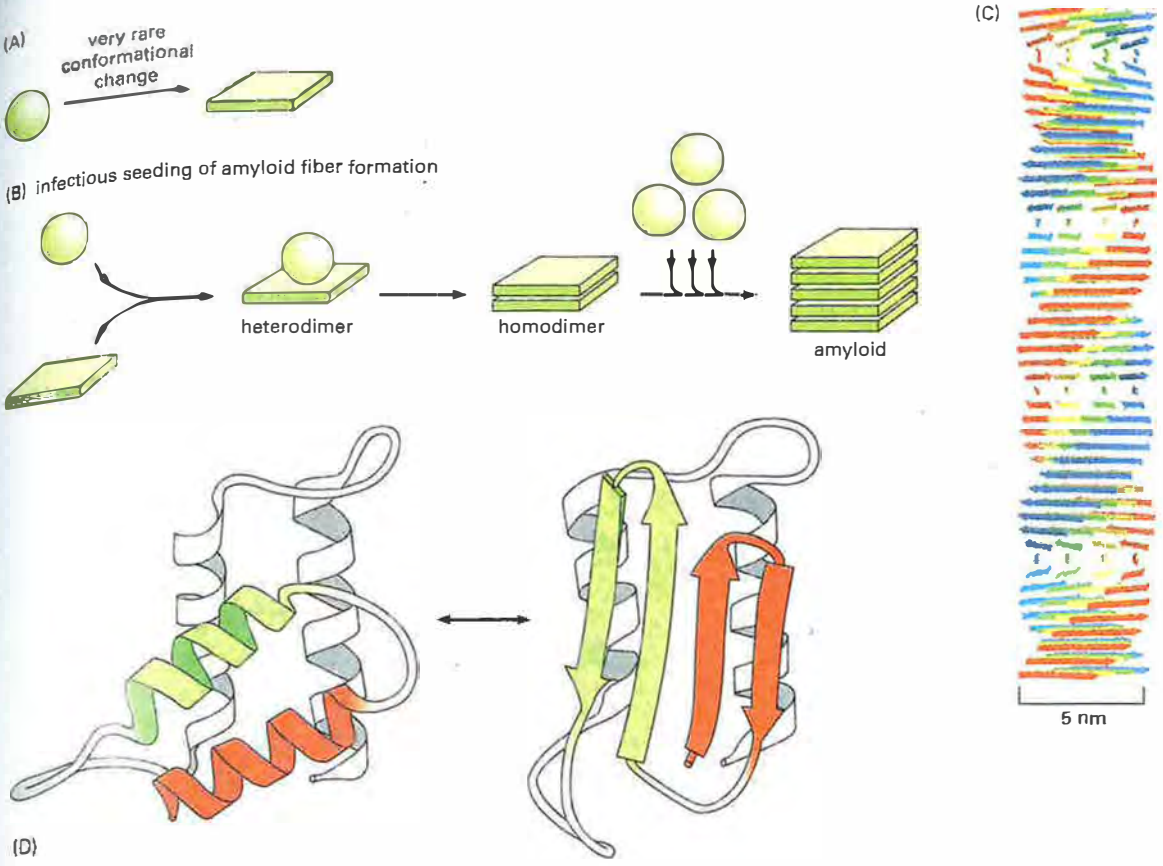


Figure 6–89 Protein aggregates that cause human disease. (A) Schematic illustration of the type of conformational change in a protein that produces material for a cross-beta filament. (B) Diagram illustrating the self-infectious nature of the protein aggregation that is central to prion diseases. PrP is highly unusual because the misfolded version of the protein, called PrP*, induces the normal PrP protein it contacts to change its conformation, as shown. Most of the human diseases caused by protein aggregation are caused by the overproduction of a variant protein that is especially prone to aggregation, but because this structure is not infectious in this way, it cannot spread from one animal to another. (C) Drawing of a cross-beta filament, a common type of protease-resistant protein aggregate found in a variety of human neurological diseases. Because the hydrogen-bond interactions in a β sheet form between polypeptide backbone atoms (see Figure 3–9), a number of different abnormally folded proteins can produce this structure. (D) One of several possible models for the conversion of PrP to PrP*, showing the likely change of two α-helices into four β-strands. Although the structure of the normal protein has been determined accurately, the structure of the infectious form is not yet known with certainty because the aggregation has prevented the use of standard structural techniques. (C, courtesy of Louise Serpell, adapted from M. Sunde et al., *J. Mol. Biol.* 273:729–739, 1997; D, adapted from S.B. Prusiner, *Trends Biochem. Sci.* 21:482–487, 1996.)

animals and humans. It can be dangerous to eat the tissues of animals that contain PrP*, as witnessed most recently by the spread of BSE (commonly referred to as the "mad cow disease") from cattle to humans in Great Britain.

Fortunately, in the absence of PrP*, PrP is extraordinarily difficult to convert to its abnormal form. Although very few proteins have the potential to misfold into an infectious conformation, a similar transformation has been discovered to be the cause of an otherwise mysterious "protein-only inheritance" observed in yeast cells.

There Are Many Steps From DNA to Protein

We have seen so far in this chapter that many different types of chemical reactions are required to produce a properly folded protein from the information contained in a gene (Figure 6–90). The final level of a properly folded protein in a cell therefore depends upon the efficiency with which each of the many steps is performed.

We discuss in Chapter 7 that cells have the ability to change the levels of their proteins according to their needs. In principle, any or all of the steps in Fig-

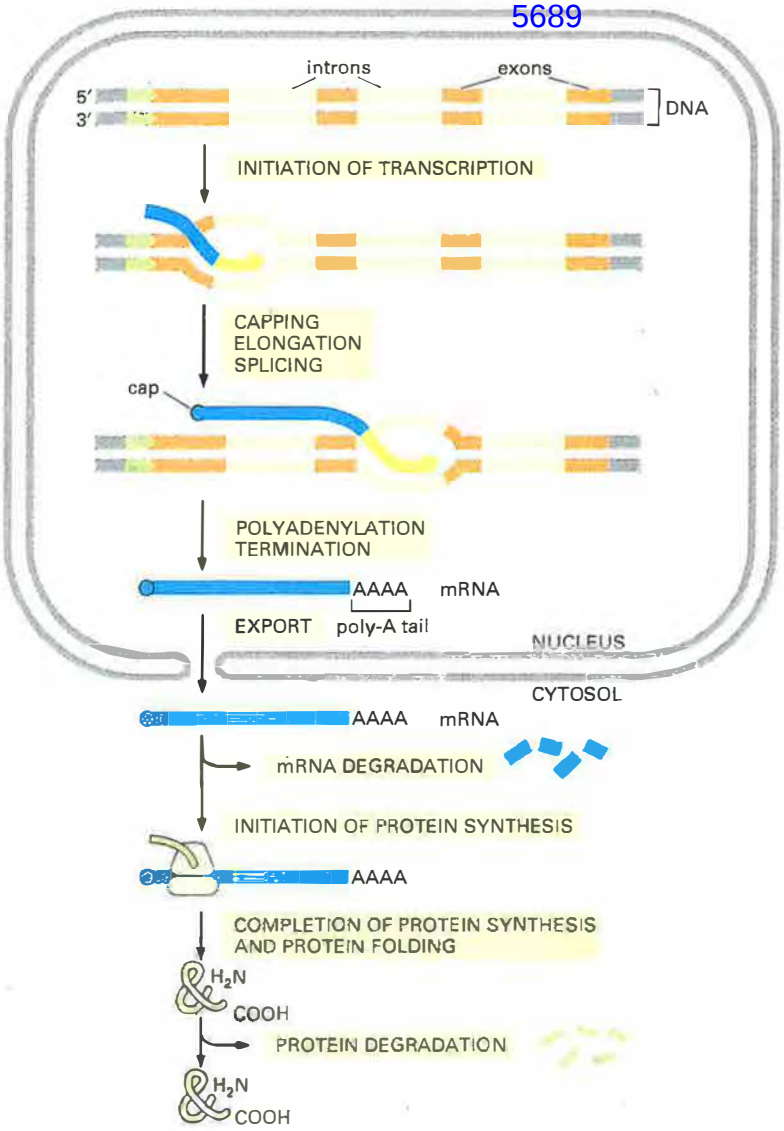


Figure 6–90 The production of a protein by a eucaryotic cell. The final level of each protein in a eucaryotic cell depends upon the efficiency of each step depicted.

ure 6–90) could be regulated by the cell for each individual protein. However, as we shall see in Chapter 7, the initiation of transcription is the most common point for a cell to regulate the expression of each of its genes. This makes sense, inasmuch as the most efficient way to keep a gene from being expressed is to block the very first step—the transcription of its DNA sequence into an RNA molecule.

Summary

The translation of the nucleotide sequence of an mRNA molecule into protein takes place in the cytoplasm on a large ribonucleoprotein assembly called a ribosome. The amino acids used for protein synthesis are first attached to a family of tRNA molecules, each of which recognizes, by complementary base-pair interactions, particular sets of three nucleotides in the mRNA (codons). The sequence of nucleotides in the mRNA is then read from one end to the other in sets of three according to the genetic code.

To initiate translation, a small ribosomal subunit binds to the mRNA molecule at a start codon (AUG) that is recognized by a unique initiator tRNA molecule. A large ribosomal subunit binds to complete the ribosome and begin the elongation phase of protein synthesis. During this phase, aminoacyl tRNAs—each bearing a specific amino acid bind sequentially to the appropriate codon in mRNA by forming complementary base pairs with the tRNA anticodon. Each amino acid is added to the C-terminal end of the growing polypeptide by means of a cycle of three sequential

steps: aminoacyl-tRNA binding, followed by peptide bond formation, followed by ribosome translocation. The mRNA molecule progresses codon by codon through the ribosome in the 5′-to-3′ direction until one of three stop codons is reached. A release factor then binds to the ribosome, terminating translation and releasing the completed polypeptide.

Eucaryotic and bacterial ribosomes are closely related, despite differences in the number and size of their rRNA and protein components. The rRNA has the dominant role in translation, determining the overall structure of the ribosome, forming the binding sites for the tRNAs, matching the tRNAs to codons in the mRNA, and providing the peptidyl transferase enzyme activity that links amino acids together during translation.

In the final steps of protein synthesis, two distinct types of molecular chaperones guide the folding of polypeptide chains. These chaperones, known as hsp60 and hsp70, recognize exposed hydrophobic patches on proteins and serve to prevent the protein aggregation that would otherwise compete with the folding of newly synthesized proteins into their correct three-dimensional conformations. This protein folding process must also compete with a highly elaborate quality control mechanism that destroys proteins with abnormally exposed hydrophobic patches. In this case, ubiquitin is covalently added to a misfolded protein by a ubiquitin ligase, and the resulting multiubiquitin chain is recognized by the cap on a proteasome to move the entire protein to the interior of the proteasome for proteolytic degradation. A closely related proteolytic mechanism, based on special degradation signals recognized by ubiquitin ligases, is used to determine the lifetime of many normally folded proteins. By this method, selected normal proteins are removed from the cell in response to specific signals.

THE RNA WORLD AND THE ORIGINS OF LIFE

To fully understand the processes occurring in present-day living cells, we need to consider how they arose in evolution. The most fundamental of all such problems is the expression of hereditary information, which today requires extraordinarily complex machinery and proceeds from DNA to protein through an RNA intermediate. How did this machinery arise? One view is that an *RNA world* existed on Earth before modern cells arose (Figure 6–91). According to this hypothesis, RNA stored both genetic information and catalyzed the chemical reactions in primitive cells. Only later in evolutionary time did DNA take over as the genetic material and proteins become the major catalyst and structural component of cells. If this idea is correct, then the transition out of the RNA world was never complete; as we have seen in this chapter, RNA still catalyzes several fundamental reactions in modern-day cells, which can be viewed as molecular fossils of an earlier world.

In this section we outline some of the arguments in support of the RNA world hypothesis. We will see that several of the more surprising features of modern-day cells, such as the ribosome and the pre-mRNA splicing machinery, are most easily explained by viewing them as descendants of a complex network of RNA-mediated interactions that dominated cell metabolism in the RNA world. We also discuss how DNA may have taken over as the genetic material, how the genetic code may have arisen, and how proteins may have eclipsed RNA to perform the bulk of biochemical catalysis in modern-day cells.

Life Requires Autocatalysis

It has been proposed that the first "biological" molecules on Earth were formed by metal-based catalysis on the crystalline surfaces of minerals. In principle, an elaborate system of molecular synthesis and breakdown (metabolism) could have existed on these surfaces long before the first cells arose. But life requires molecules that possess a crucial property: the ability to catalyze reactions that lead, directly or indirectly, to the production of more molecules like themselves. Catalysts with this special self-promoting property can use raw materials to

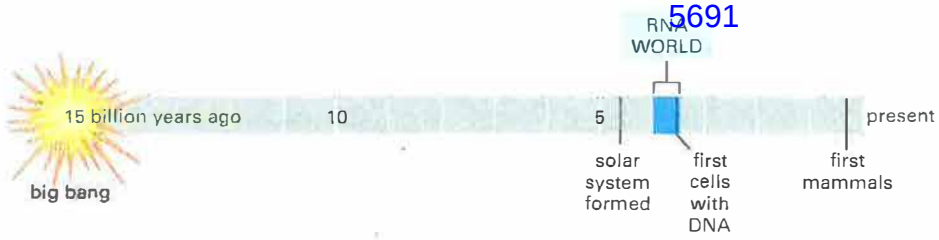


Figure 6–91 Time line for the universe, suggesting the early existence of an RNA world of living systems.

reproduce themselves and thereby divert these same materials from the production of other substances. But what molecules could have had such autocatalytic properties in early cells? In present-day cells the most versatile catalysts are polypeptides, composed of many different amino acids with chemically diverse side chains and, consequently, able to adopt diverse three-dimensional forms that bristle with reactive chemical groups. But, although polypeptides are versatile as catalysts, there is no known way in which one such molecule can reproduce itself by directly specifying the formation of another of precisely the same sequence.

Polynucleotides Can Both Store Information and Catalyze Chemical Reactions

Polynucleotides have one property that contrasts with those of polypeptides: they can directly guide the formation of exact copies of their own sequence. This capacity depends on complementary base pairing of nucleotide subunits, which enables one polynucleotide to act as a template for the formation of another. As we have seen in this and the preceding chapter, such complementary templating mechanisms lie at the heart of DNA replication and transcription in modern-day cells.

But the efficient synthesis of polynucleotides by such complementary templating mechanisms requires catalysts to promote the polymerization reaction: without catalysts, polymer formation is slow, error-prone, and inefficient. Today, template-based nucleotide polymerization is rapidly catalyzed by protein enzymes—such as the DNA and RNA polymerases. How could it be catalyzed before proteins with the appropriate enzymatic specificity existed? The beginnings of an answer to this question were obtained in 1982, when it was discovered that RNA molecules themselves can act as catalysts. We have seen in this chapter, for example, that a molecule of RNA is the catalyst for the peptidyl transferase reaction that takes place on the ribosome. The unique potential of RNA molecules to act both as information carrier and as catalyst forms the basis of the RNA world hypothesis.

RNA therefore has all the properties required of a molecule that could catalyze its own synthesis (Figure 6–92). Although self-replicating systems of RNA molecules have not been found in nature, scientists are hopeful that they can be constructed in the laboratory. While this demonstration would not prove that self-replicating RNA molecules were essential in the origin of life on Earth, it would certainly suggest that such a scenario is possible.

A Pre-RNA World Probably Predates the RNA World

Although RNA seems well suited to form the basis for a self-replicating set of biochemical catalysts, it is unlikely that RNA was the first kind of molecule to do so.

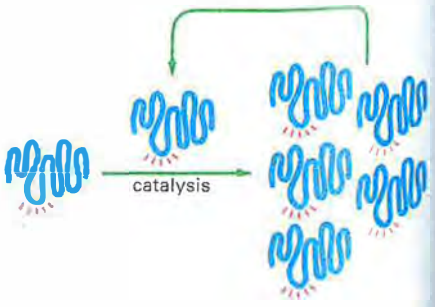


Figure 6–92 An RNA molecule that can catalyze its own synthesis. This hypothetical process would require catalysis of the production of both a second RNA strand of complementary nucleotide sequence and the use of this second RNA molecule as a template to form many molecules of RNA with the original sequence. The *red* rays represent the active site of this hypothetical RNA enzyme.

From a purely chemical standpoint, it is difficult to imagine how long RNA molecules could be formed initially by purely nonenzymatic means. For one thing, the precursors of RNA, the ribonucleotides, are difficult to form nonenzymatically. Moreover, the formation of RNA requires that a long series of 3′ to 5′ phosphodiester linkages form in the face of a set of competing reactions, including hydrolysis, 2′ to 5′ linkages, 5′ to 5′ linkages, and so on. Given these problems, it has been suggested that the first molecules to possess both catalytic activity and information storage capabilities may have been polymers that resemble RNA but are chemically simpler (Figure 6–93). We do not have any remnants of these compounds in present-day cells, nor do such compounds leave fossil records. Nonetheless, the relative simplicity of these "RNA-like polymers" make them better candidates than RNA itself for the first biopolymers on Earth that had both information storage capacity and catalytic activity.

The transition between the pre-RNA world and the RNA world would have occurred through the synthesis of RNA using one of these simpler compounds as both template and catalyst. The plausibility of this scheme is supported by laboratory experiments showing that one of these simpler forms (PNA—see Figure 6–93) can act as a template for the synthesis of complementary RNA molecules, because the overall geometry of the bases is similar in the two molecules. Presumably, pre-RNA polymers also catalyzed the formation of ribonucleotide precursors from simpler molecules. Once the first RNA molecules had been produced, they could have diversified gradually to take over the functions originally carried out by the pre-RNA polymers, leading eventually to the postulated RNA world.

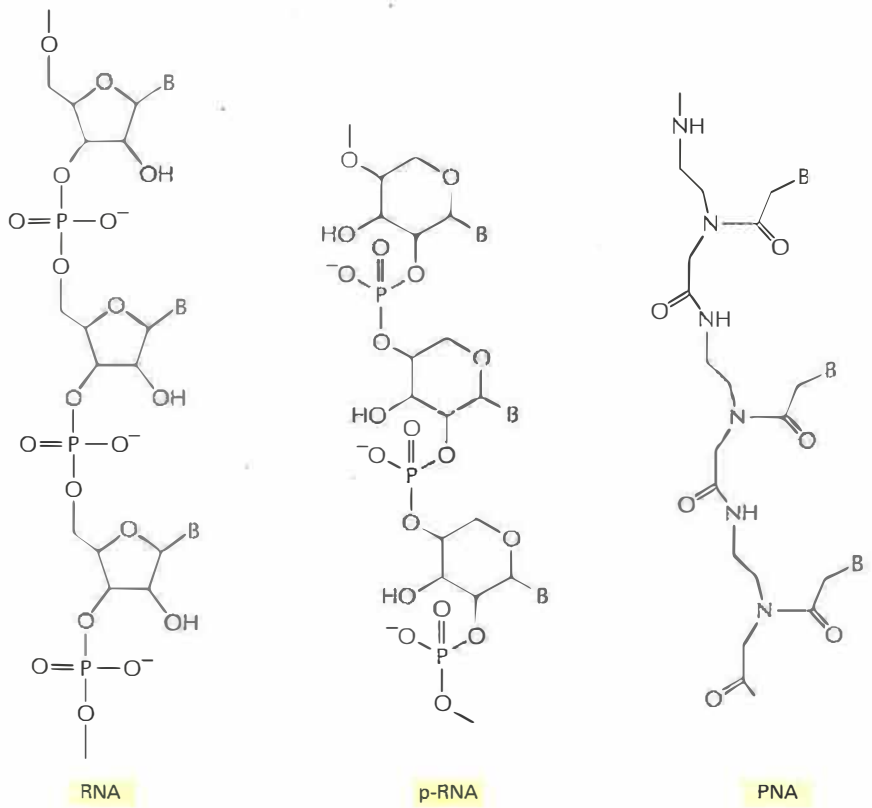


Figure 6–93 Structures of RNA and two related information-carrying polymers. In each case, B indicates the positions of purine and pyrimidine bases. The polymer p-RNA (pyranosyl-RNA) is RNA in which the furanose (five-membered ring) form of ribose has been replaced by the pyranose (six-membered ring) form. In PNA (peptide nucleic acid), the ribose phosphate backbone of RNA has been replaced by the peptide backbone found in proteins. Like RNA, both p-RNA and PNA can form double helices through complementary base-pairing, and each could therefore in principle serve as a template for its own synthesis (see Figure 6–92).

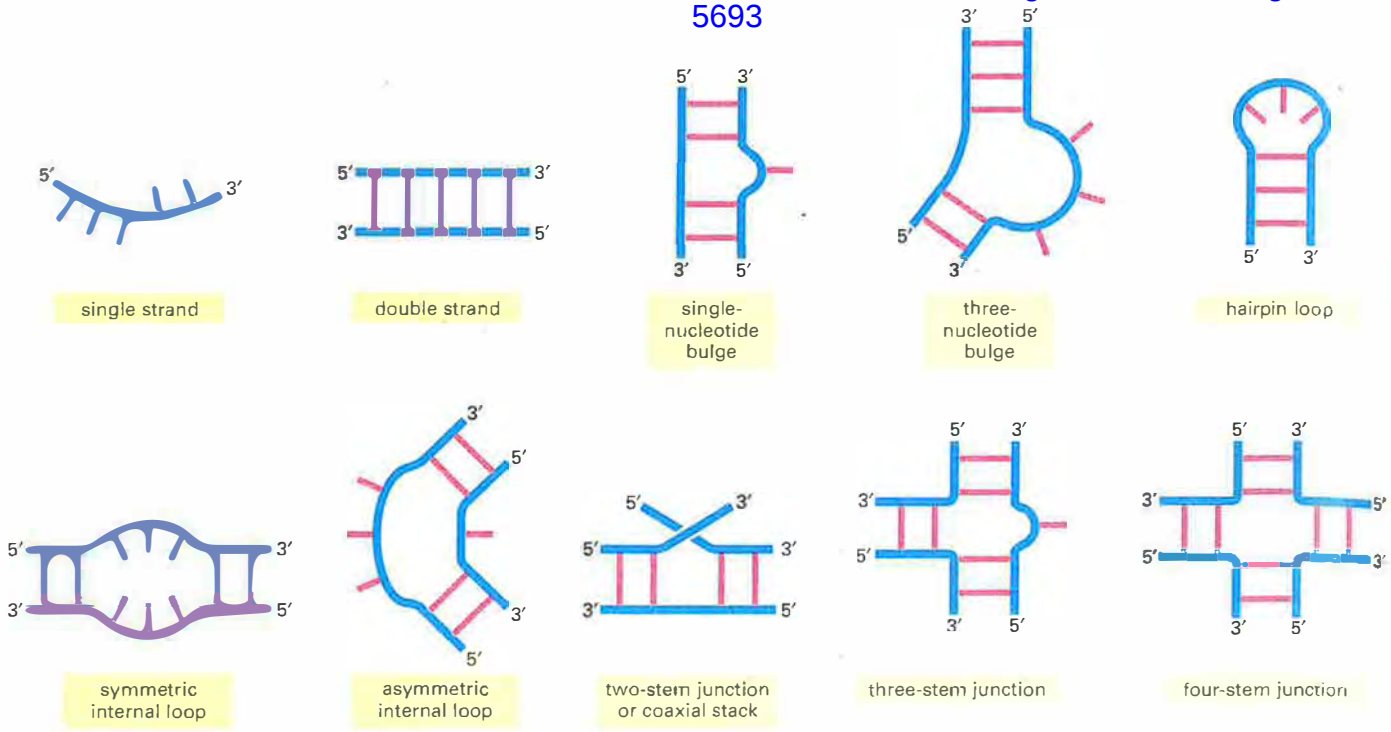


Figure 6–94 Common elements of RNA secondary structure. Conventional, complementary base-pairing interactions are indicated by *red* "rungs" in double-helical portions of the RNA.

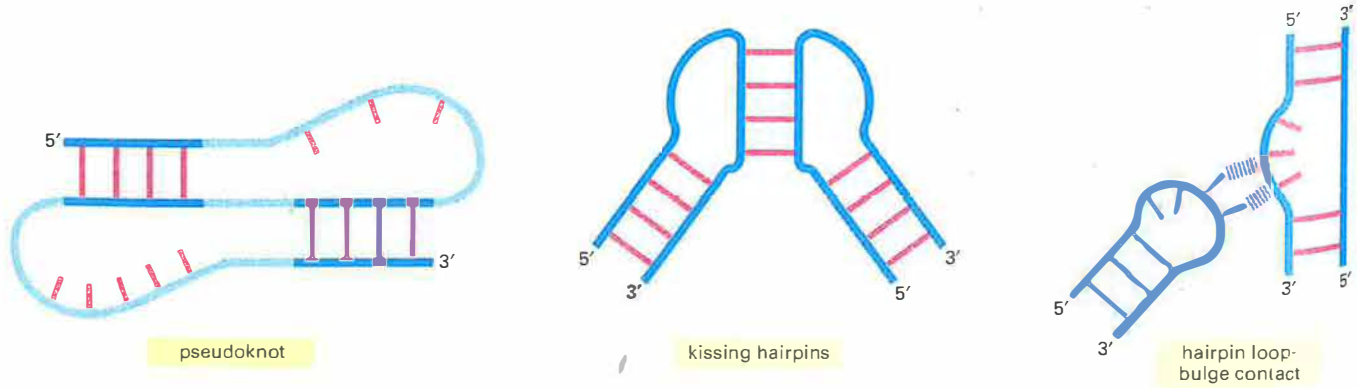
Single-stranded RNA Molecules Can Fold into Highly Elaborate Structures

We have seen that complementary base-pairing and other types of hydrogen bonds can occur between nucleotides in the same chain, causing an RNA molecule to fold up in a unique way determined by its nucleotide sequence (see, for example, Figures 6–6, 6–52, and 6–67). Comparisons of many RNA structures have revealed conserved motifs, short structural elements that are used over and over again as parts of larger structures. Some of these RNA secondary structural motifs are illustrated in Figure 6–94. In addition, a few common examples of more complex and often longer-range interactions, known as RNA tertiary interactions, are shown in Figure 6–95.

Protein catalysts require a surface with unique contours and chemical properties on which a given set of substrates can react (discussed in Chapter 3). In exactly the same way, an RNA molecule with an appropriately folded shape can serve as an enzyme (Figure 6–96). Like some proteins, many of these ribozymes work by positioning metal ions at their active sites. This feature gives them a wider range of catalytic activities than can be accounted for solely by the limited chemical groups of the polynucleotide chain.

Relatively few catalytic RNAs exist in modern-day cells, however, and much of our inference about the RNA world has come from experiments in which large pools of RNA molecules of random nucleotide sequences are generated in the laboratory. Those rare RNA molecules with a property specified by the experimenter are then selected out and studied (Figure 6–97). Experiments of this type

Figure 6–95 Examples of RNA tertiary interactions. Some of these interactions can join distant parts of the same RNA molecule or bring two separate RNA molecules together.



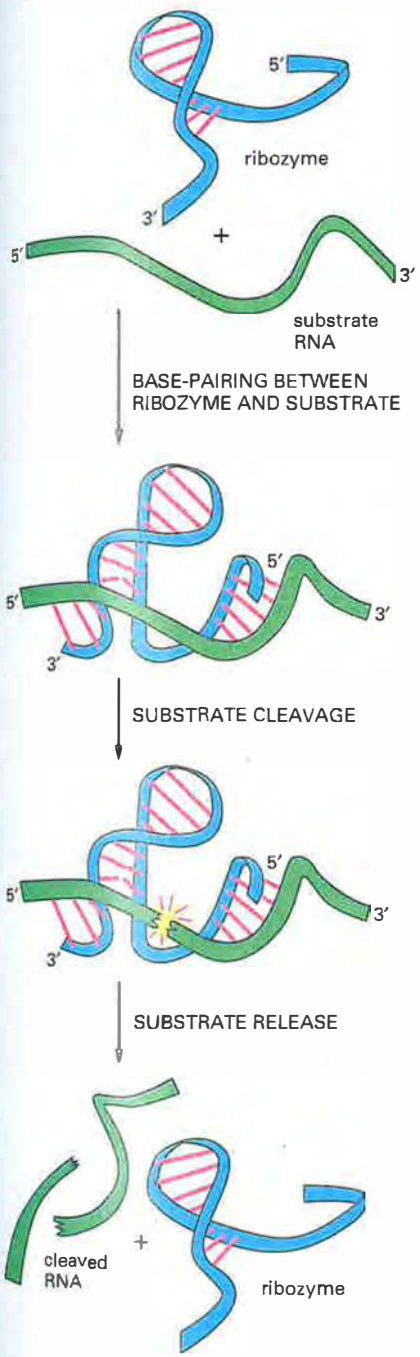


Figure 6–96 *(above)* A ribozyme. This simple RNA molecule catalyzes the cleavage of a second RNA at a specific site. This ribozyme is found embedded in larger RNA genomes—called viroids—which infect plants. The cleavage, which occurs in nature at a distant location on the same RNA molecule that contains the ribozyme, is a step in the replication of the viroid genome. Although not shown in the figure, the reaction requires a molecule of Mg positioned at the active site. (Adapted from T.R. Cech and O.C. Uhlenbeck, *Nature* 372:39–40, 1994.)

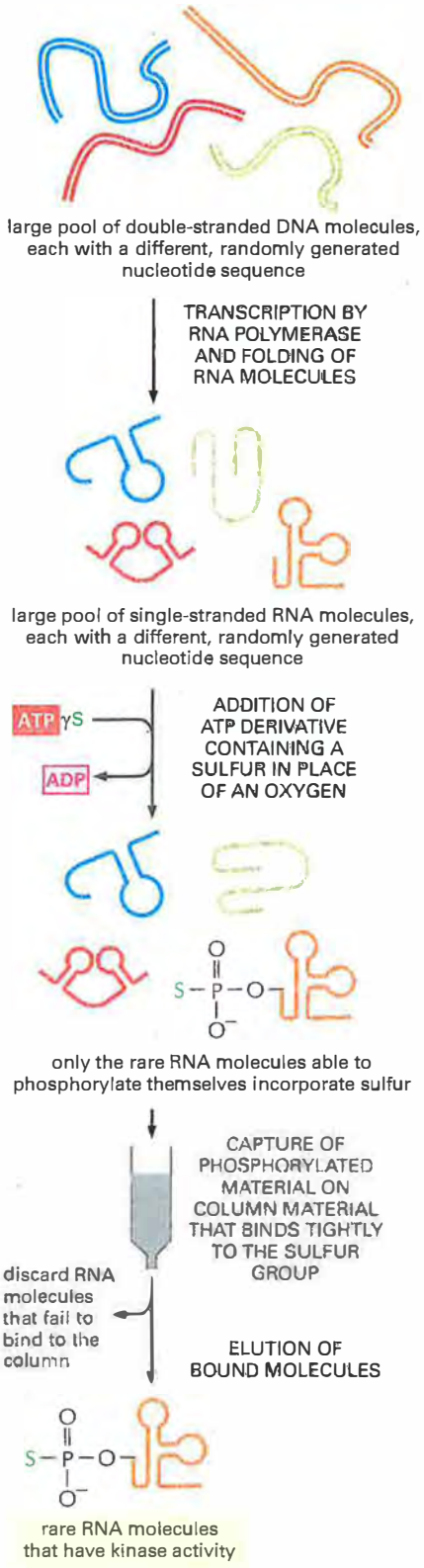


Figure 6–97 *(left)* *In vitro* selection of a synthetic ribozyme. Beginning with a large pool of nucleic acid molecules synthesized in the laboratory, those rare RNA molecules that possess a specified catalytic activity can be isolated and studied. Although a specific example (that of an autophosphorylating ribozyme) is shown, variations of this procedure have been used to generate many of the ribozymes listed in Table 6–4. During the autophosphorylation step, the RNA molecules are sufficiently dilute to prevent the "cross"-phosphorylation of additional RNA molecules. In reality, several repetitions of this procedure are necessary to select the very rare RNA molecules with catalytic activity. Thus the material initially eluted from the column is converted back into DNA, amplified many fold (using reverse transcriptase and PCR as explained in Chapter 8), transcribed back into RNA, and subjected to repeated rounds of selection. (Adapted from J.R. Lorsch and J.W. Szostak, *Nature* 371:31–36, 1994.)

TABLE 6–4 Some Biochemical Reactions That Can Be Catalyzed by Ribozymes

ACTIVITY	RIBOZYMES
Peptide bond formation in protein synthesis	ribosomal RNA
RNA cleavage, RNA ligation	self-splicing RNAs; also *in vitro* selected RNA
DNA cleavage	self-splicing RNAs
RNA splicing	self-splicing RNAs, perhaps RNAs of the spliceosome
RNA polymerizaton	*in vitro* selected RNA
RNA and DNA phosphorylation	*in vitro* selected RNA
RNA aminoacylation	*in vitro* selected RNA
RNA alkylation	*in vitro* selected RNA
Amide bond formation	*in vitro* selected RNA
Amide bond cleavage	*in vitro* selected RNA
Glycosidic bond formation	*in vitro* selected RNA
Porphyrin metalation	*in vitro* selected RNA

have created RNAs that can catalyze a wide variety of biochemical reactions (Table 6–4), and suggest that the main difference between protein enzymes and ribozymes lies in their maximum reaction speed, rather than in the diversity of the reactions that they can catalyze.

Like proteins, RNAs can undergo allosteric conformational changes, either in response to small molecules or to other RNAs. One artificially created ribozyme can exist in two entirely different conformations, each with a different catalytic activity (Figure 6–98). Moreover, the structure and function of the rRNAs in the ribosome alone have made it clear that RNA is an enormously versatile molecule. It is therefore easy to imagine that an RNA world could reach a high level of biochemical sophistication.

Self-Replicating Molecules Undergo Natural Selection

The three-dimensional folded structure of a polynucleotide affects its stability, its actions on other molecules, and its ability to replicate. Therefore, certain polynucleotides will be especially successful in any self-replicating mixture. Because errors inevitably occur in any copying process, new variant sequences of these polynucleotides will be generated over time.

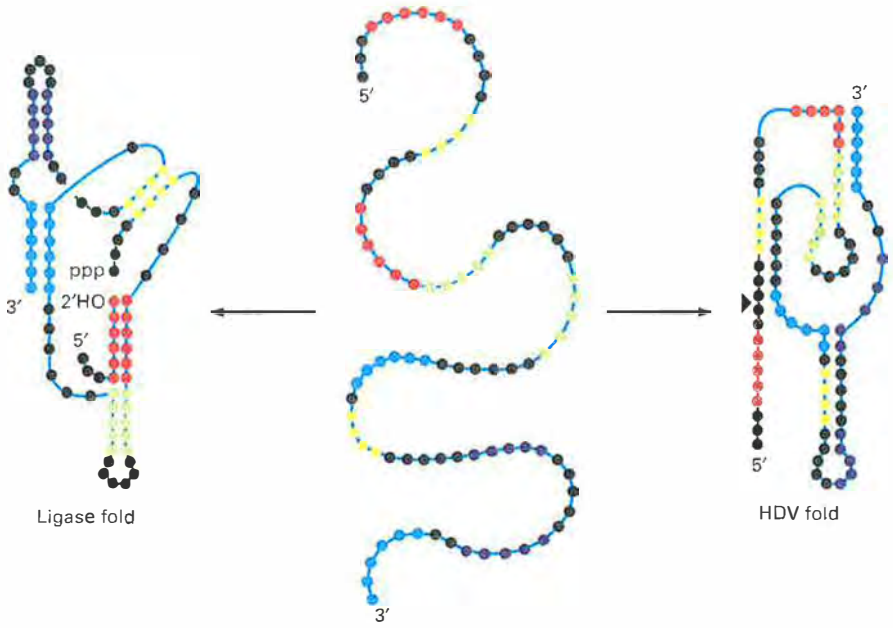


Figure 6–98 An RNA molecule that folds into two different ribozymes. This 88-nucleotide RNA, created in the laboratory, can fold into a ribozyme that carries out a self-ligation reaction *(left)* or a self-cleavage reaction *(right)*. The ligation reaction forms a 2′,5′ phosphodiester linkage with the release of pyrophosphate. This reaction seals the gap *(gray* shading), which was experimentally introduced into the RNA molecule. In the reaction carried out by the HDV fold, the RNA is cleaved at this same position, indicated by the *arrowhead*. This cleavage resembles that used in the life cycle of HDV, a hepatitis B satellite virus, hence the name of the fold. Each nucleotide is represented by a *colored dot*, with the colors used simply to clarify the two different folding patterns. The folded structures illustrate the secondary structures of the two ribozymes with regions of base-pairing indicated by close oppositions of the *colored dots*. Note that the two ribozyme folds have no secondary structure in common. (Adapted from E.A. Schultes and D.P. Bartel, *Science* 289:448–452, 2000.)

Certain catalytic activities would have had a cardinal importance in the early evolution of life. Consider in particular an RNA molecule that helps to catalyze the process of templated polymerization, taking any given RNA molecule as a template. (This ribozyme activity has been directly demonstrated *in vitro*, albeit in a rudimentary form that can only synthesize moderate lengths of RNA.) Such a molecule, by acting on copies of itself, can replicate. At the same time, it can promote the replication of other types of RNA molecules in its neighborhood (Figure 6–99). If some of these neighboring RNAs have catalytic actions that help the survival of RNA in other ways (catalyzing ribonucleotide production, for example), a set of different types of RNA molecules, each specialized for a different activity, may evolve into a cooperative system that replicates with unusually great efficiency.

One of the crucial events leading to the formation of effective self-replicating systems must have been the development of individual compartments. For example, a set of mutually beneficial RNAs (such as those of Figure 6–99) could replicate themselves only if all the RNAs were to remain in the neighborhood of the RNA that is specialized for templated polymerization. Moreover, if these RNAs were free to diffuse among a large population of other RNA molecules, they could be co-opted by other replicating systems, which would then compete with the original RNA system for raw materials. Selection of a set of RNA molecules according to the quality of the self-replicating systems they generated could not occur efficiently until some form of compartment evolved to contain them and thereby make them available only to the RNA that had generated them. An early, crude form of compartmentalization may have been simple adsorption on surfaces or particles.

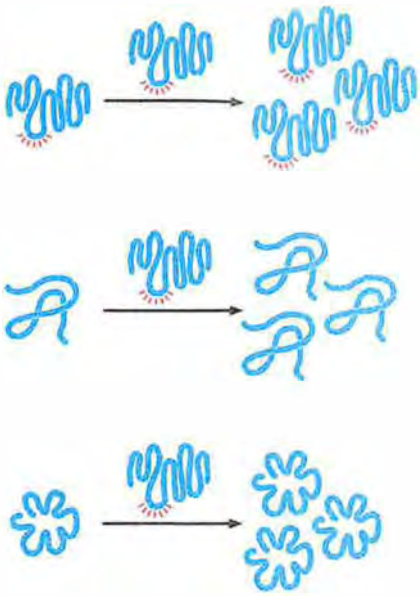
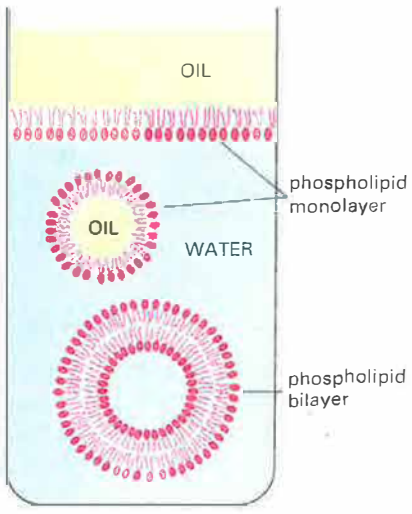


Figure 6–99 A family of mutually supportive RNA molecules, one catalyzing the reproduction of the others.

The need for more sophisticated types of containment is easily fulfilled by a class of small molecules that has the simple physicochemical property of being *amphipathic*, that is, consisting of one part that is hydrophobic (water insoluble) and another part that is hydrophilic (water soluble). When such molecules are placed in water they aggregate, arranging their hydrophobic portions as much in contact with one another as possible and their hydrophilic portions in contact with the water. Amphipathic molecules of appropriate shape spontaneously aggregate to form *bilayers*, creating small closed vesicles whose aqueous contents are isolated from the external medium (Figure 6–100). The phenomenon can be demonstrated in a test tube by simply mixing phospholipids and water together: under appropriate conditions, small vesicles will form. All present-day cells are surrounded by a *plasma membrane* consisting of amphipathic molecules—mainly phospholipids—in this configuration; we discuss these molecules in detail in Chapter 10.

Presumably, the first membrane-bounded cells were formed by the spontaneous assembly of a set of amphipathic molecules, enclosing a self-replicating mixture of RNA (or pre-RNA) and other molecules. It is not clear at what point in the evolution of biological catalysts this first occurred. In any case, once RNA molecules were sealed within a closed membrane, they could begin to evolve in earnest as carriers of genetic instructions: they could be selected not merely on the basis of their own structure, but also according to their effect on the other molecules in the same compartment. The nucleotide sequences of the RNA molecules could now be expressed in the character of a unitary living cell.



How Did Protein Synthesis Evolve?

The molecular processes underlying protein synthesis in present-day cells seem inextricably complex. Although we understand most of them, they do not make conceptual sense in the way that DNA transcription, DNA repair, and DNA

Figure 6–100 Formation of membrane by phospholipids. Because these molecules have hydrophilic heads and lipophilic tails, they align themselves at an oil/water interface with their heads in the water and their tails in the oil. In the water they associate to form closed bilayer vesicles in which the lipophilic tails are in contact with one another and the hydrophilic heads are exposed to the water.

replication do. It is especially difficult to imagine how protein synthesis evolved because it is now performed by a complex interlocking system of protein and RNA molecules; obviously the proteins could not have existed until an early version of the translation apparatus was already in place. Although we can only speculate on the origins of protein synthesis and the genetic code, several experimental approaches have provided possible scenarios.

In vitro RNA selection experiments of the type summarized previously in Figure 6–97 have produced RNA molecules that can bind tightly to amino acids. The nucleotide sequences of these RNAs often contain a disproportionately high frequency of codons for the amino acid that is recognized. For example, RNA molecules that bind selectively to arginine have a preponderance of Arg codons and those that bind tyrosine have a preponderance of Tyr codons. This correlation is not perfect for all the amino acids, and its interpretation is controversial, but it raises the possibility that a limited genetic code could have arisen from the direct association of amino acids with specific sequences of RNA, with RNAs serving as a crude template to direct the non-random polymerization of a few different amino acids. In the RNA world described previously, any RNA that helped guide the synthesis of a useful polypeptide would have a great advantage in the evolutionary struggle for survival.

In present-day cells, tRNA adaptors are used to match amino acids to codons, and proteins catalyze tRNA aminoacylation. However, ribozymes created in the laboratory can perform specific tRNA aminoacylation reactions, so it is plausible that tRNA-like adaptors could have arisen in an RNA world. This development would have made the matching of "mRNA" sequences to amino acids more efficient, and it perhaps allowed an increase in the number of amino acids that could be used in templated protein synthesis.

Finally, the efficiency of early forms of protein synthesis would be increased dramatically by the catalysis of peptide bond formation. This evolutionary development presents no conceptual problem since, as we have seen, this reaction is catalyzed by rRNA in present-day cells. One can envision a crude peptidyl transferase ribozyme, which, over time, grew larger and acquired the ability to position charged tRNAs accurately on RNA templates—leading eventually to the modern ribosome. Once protein synthesis evolved, the transition to a protein-dominated world could proceed, with proteins eventually taking over the majority of catalytic and structural tasks because of their greater versatility, with 20 rather than 4 different subunits.

All Present-day Cells Use DNA as Their Hereditary Material

The cells of the RNA world would presumably have been much less complex and less efficient in reproducing themselves than even the simplest present-day cells, since catalysis by RNA molecules is less efficient than that by proteins. They would have consisted of little more than a simple membrane enclosing a set of self-replicating molecules and a few other components required to provide the materials and energy for their replication. If the evolutionary speculations about RNA outlined above are correct, these early cells would also have differed fundamentally from the cells we know today in having their hereditary information stored in RNA rather than in DNA (Figure 6–101).

Evidence that RNA arose before DNA in evolution can be found in the chemical differences between them. Ribose, like glucose and other simple carbohydrates, can be formed from formaldehyde (HCHO), a simple chemical which is readily produced in laboratory experiments that attempt to simulate conditions on the primitive Earth. The sugar deoxyribose is harder to make, and in present-day cells it is produced from ribose in a reaction catalyzed by a protein enzyme, suggesting that ribose predates deoxyribose in cells. Presumably, DNA appeared on the scene later, but then proved more suitable than RNA as a permanent repository of genetic information. In particular, the deoxyribose in its sugar-phosphate backbone makes chains of DNA chemically more stable than chains of RNA, so that much greater lengths of DNA can be maintained without breakage.

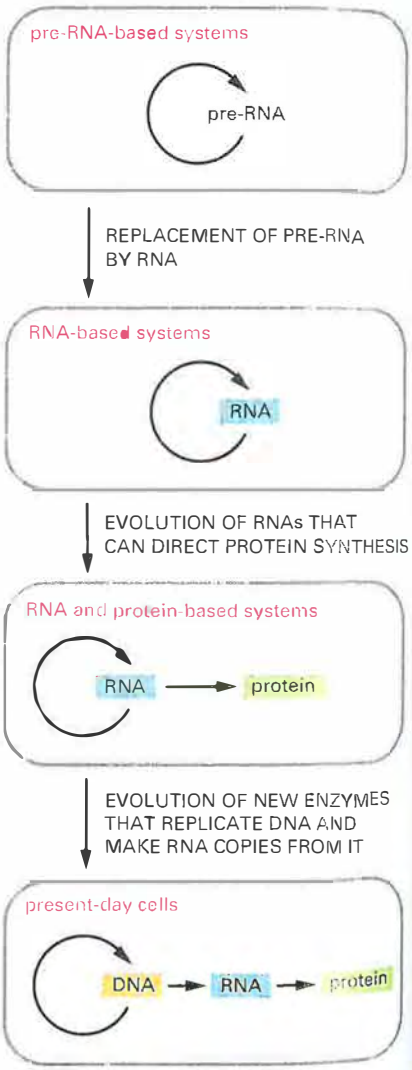


Figure 6–101 The hypothesis that RNA preceded DNA and proteins in evolution. In the earliest cells, pre-RNA molecules would have had combined genetic, structural, and catalytic functions and these functions would have gradually been replaced by RNA. In present-day cells, DNA is the repository of genetic information, and proteins perform the vast majority of catalytic functions in cells. RNA primarily functions today as a go-between in protein synthesis, although it remains a catalyst for a number of crucial reactions.

The other differences between RNA and DNA—the double-helical structure of DNA and the use of thymine rather than uracil—further enhance DNA stability by making the many unavoidable accidents that occur to the molecule much easier to repair, as discussed in detail in Chapter 5 (see pp. 269–272).

Summary

From our knowledge of present-day organisms and the molecules they contain, it seems likely that the development of the directly autocatalytic mechanisms fundamental to living systems began with the evolution of families of molecules that could catalyze their own replication. With time, a family of cooperating RNA catalysts probably developed the ability to direct synthesis of polypeptides. DNA is likely to have been a late addition: as the accumulation of additional protein catalysts allowed more efficient and complex cells to evolve, the DNA double helix replaced RNA as a more stable molecule for storing the increased amounts of genetic information required by such cells.

References

General

Gesteland RF, Cech TR & Atkins JF (eds) (1999) The RNA World, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Hartwell L, Hood L, Goldberg ML et al. (2000) Genetics: from Genes to Genomes. Boston: McGraw Hill.

Lewin B (2000) Genes VII. Oxford: Oxford University Press.

Lodish H, Berk A, Zipursky SL et al. (2000) Molecular Cell Biology, 4th edn. New York: WH Freeman.

Stent GS (1971) Molecular Genetics: An Introductory Narrative. San Francisco: WH Freeman.

Stryer L (1995) Biochemistry, 4th edn. New York: WH Freeman.

Watson JD, Hopkins NH, Roberts JW et al. (1987) Molecular Biology of the Gene, 4th edn. Menlo Park, CA: Benjamin/Cummings.

From DNA to RNA

Berget SM, Moore C & Sharp PA (1977) Spliced segments at the 5′ terminus of adenovirus 2 late mRNA. *Proc. Natl. Acad. Sci. USA* 74, 3171–3175.

Black DL (2000) Protein diversity from alternative splicing: a challenge for bioinformatics and post-genome biology. *Cell* 103, 367–370.

Brenner S, Jacob F & Meselson M (1961) An unstable intermediate carrying information from genes to ribosomes for protein synthesis. *Nature* 190, 576–581.

Cech TR (1990) Nobel lecture. Self-splicing and enzymatic activity of an intervening sequence RNA from Tetrahymena. *Biosci. Rep.* 10, 239–261.

Chow LT, Gelinas RE, Broker TR et al. (1977) An amazing sequence arrangement at the 5′ ends of adenovirus 2 messenger RNA. *Cell* 12, 1–8.

Conaway JW, Shilatifard A, Dvir A & Conaway RC (2000) Control of elongation by RNA polymerase II. *Trends Biochem. Sci.* 25, 375–380.

Cramer P, Bushnell DA, Fu J et al. (2000) Architecture of RNA polymerase II and implications for the transcription mechanism. *Science* 288, 640–649.

Crick F (1979) Split genes and RNA splicing. *Science* 204, 264–271.

Daneholt B (1997) A look at messenger RNP moving through the nuclear pore. *Cell* 88, 585–588.

Darnell JE, Jr. (1985) RNA. *Sci. Am.* 253(4), 68–78.

Dvir A, Conaway JW & Conaway RC (2001) Mechanism of transcription initiation and promoter escape by RNA polymerase II. *Curr. Opin. Genet. Dev.* 11, 209–214.

Ebright RH (2000) RNA polymerase: structural similarities between bacterial RNA polymerase and eukaryotic RNA polymerase II. *J. Mol. Biol.* 304, 687–698.

Eddy SR (1999) Noncoding RNA genes. *Curr. Opin. Genet. Dev.* 9, 695–699.

Green MR (2000) TBP-associated factors (TAFIIs): multiple, selective transcriptional mediators in common complexes. *Trends Biochem. Sci.* 25, 59–63.

Harley CB & Reynolds RP (1987) Analysis of E. coli promoter sequences. *Nucleic Acids Res.* 15, 2343–2361.

Hirose Y & Manley JL (2000) RNA polymerase II and the integration of nuclear events. *Genes Dev.* 14, 1415–1429.

Kadonaga JT (1998) Eukaryotic transcription: an interlaced network of transcription factors and chromatin-modifying machines. *Cell* 92, 307–313.

Lewis JD & Tollervey D (2000) Like attracts like: getting RNA processing together in the nucleus. *Science* 288, 1385–1389.

Lisser S & Margalit H (1993) Compilation of E. coli mRNA promoter sequences. *Nucleic Acids Res.* 21, 1507–1516.

Minvielle-Sebastia L & Keller W (1999) mRNA polyadenylation and its coupling to other RNA processing reactions and to transcription. *Curr. Opin. Cell Biol.* 11, 352–357.

Mooney RA & Landick R (1999) RNA polymerase unveiled. *Cell* 98, 687–690.

Olson MO, Dundr M & Szebeni A (2000) The nucleolus: an old factory with unexpected capabilities. *Trends Cell Biol.* 10, 189–196.

Proudfoot N (2000) Connecting transcription to messenger RNA processing. *Trends Biochem. Sci.* 25, 290–293.

Reed R (2000) Mechanisms of fidelity in pre-mRNA splicing. *Curr. Opin. Cell Biol.* 12, 340–345.

Roeder RG (1996) The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem. Sci.* 21, 327–335.

Shatkin AJ & Manley JL (2000) The ends of the affair: capping and polyadenylation. *Nat. Struct. Biol.* 7, 838–842.

Smith CW & Valcarcel J (2000) Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends Biochem. Sci.* 25, 381–388.

Staley JP & Guthrie C (1998) Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell* 92, 315–326.

Tarn WY & Steitz JA (1997) Pre-mRNA splicing: the discovery of a new spliceosome doubles the challenge. *Trends Biochem. Sci.* 22, 132–137.

von Hippel PH (1998) An integrated model of the transcription complex in elongation, termination, and editing. *Science* 281, 660–665.

From RNA to Protein

Abelson J, Trotta CR & Li H (1998) tRNA splicing. *J. Biol. Chem.* 273, 12685–12688.

Anfinsen CB (1973) Principles that govern the folding of protein chains. *Science* 181, 223–230.

Cohen FE (1999) Protein misfolding and prion diseases. *J. Mol. Biol.* 293, 313–320.

Crick FHC (1966) The genetic code: III. *Sci. Am.* 215(4), 55–62.

Fedorov AN & Baldwin TO (1997) Cotranslational protein folding. *J. Biol. Chem.* 272, 32715–32718.

Frank J (2000) The ribosome—a macromolecular machine par excellence. *Chem. Biol.* 7, R133–141.

Green R (2000) Ribosomal translocation: EF-G turns the crank. *Curr. Biol.* 10, R369–373.

Hartl FU (1996) Molecular chaperones in cellular protein folding. *Nature* 381, 571–580.

Hershko A, Ciechanover A & Varshavsky A (2000) The ubiquitin system. *Nat. Med.* 6, 1073–1081.

Ibba M & Soll D (2000) Aminoacyl-tRNA synthesis. *Annu. Rev. Biochem.* 69, 617–650.

Kozak M (1999) Initiation of translation in prokaryotes and eukaryotes. *Gene* 234, 187–208.

Nissen P, Hansen J, Ban N et al. (2000) The structural basis of ribosome activity in peptide bond synthesis. *Science* 289, 920–930.

Nureki O, Vassylyev DG, Tateno M et al. (1998) Enzyme structure with two catalytic sites for double-sieve selection of substrate. *Science* 280, 578–582.

Prusiner SB (1998) Nobel lecture. Prions. *Proc. Natl. Acad. Sci. USA* 95, 13363–13383.

Rich A & Kim SH (1978) The three-dimensional structure of transfer RNA. *Sci. Am.* 238(1), 52–62.

Sachs AB & Varani G (2000) Eukaryotic translation initiation: there are (at least) two side to every story. *Nat. Struct. Biol.* 7, 356–361.

The Genetic Code. (1966) Cold Spring Harbor Symposium on Quantitative Biology, vol XXXI. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Turner GC & Varshavsky A (2000) Detecting and measuring cotranslational protein degradation *in vivo*. *Science* 289, 2117–2120.

Varshavsky A, Turner G, Du F et al. (2000) The ubiquitin system and the N-end rule pathway. *Biol. Chem.* 381, 779–789.

Voges D, Zwickl P & Baumeister W (1999) The 26S proteasome: a molecular machine designed for controlled proteolysis. *Annu. Rev. Biochem.* 68, 1015–1068.

Wilson KS & Noller HF (1998) Molecular movement inside the translational engine. *Cell* 92, 337–349.

Wimberly BT, Brodersen DE, Clemons WM et al. (2000) Structure of the 30S ribosomal subunit. *Nature* 407, 327–339.

Yusupov MM, Yusupova GZ, Baucom A et al. (2001) Crystal structure of the ribosome at 5.5A resolution. *Science* 292, 883–896.

The RNA World and the Origins of Life

Bartel DP & Unrau PJ (1999) Constructing an RNA world. *Trends Cell Biol.* 9, M9–M13.

Joyce GF (1992) Directed molecular evolution. *Sci. Am.* 267(6), 90–97.

Knight RD & Landweber LF (2000) The early evolution of the genetic code. *Cell* 101, 569–572.

Orgel L (2000) Origin of life. A simpler nucleic acid. *Science* 290, 1306–1307.

Szathmary E (1999) The origin of the genetic code: amino acids as cofactors in an RNA world. *Trends Genet.* 15, 223–229.

EXHIBIT 8

Psychosocial Aspects of Death and Dying in Duchenne Muscular Dystrophy

Julie G. Botvin Madorsky, MD, Linda M. Radford, MA, Elizabeth M. Neumann, PhD
Casa Colina Hospital for Rehabilitative Medicine, Pomoma, CA 91767

ABSTRACT. Madorsky JGB, Radford LM, Neumann EM: Psychosocial aspects of death and dying in Duchenne muscular dystrophy. Arch Phys Med Rehabil 65:79–82, 1984.

• An exploratory survey of need and availability of life expectancy information and counseling was conducted among 32 patients who have Duchenne muscular dystrophy, 75 parents, and 43 professional health care workers in the muscular dystrophy clinics of the greater Los Angeles area. Regardless of their ages, which ranged from 12 years to 21, the majority of the patients, as well as the majority of the parents, rated life expectancy information and counseling as an important, but relatively unmet, need. Professionals perceived these services as more available than did the patients and their parents. Implications for intervention include restructuring psychosocial services for the entire family and providing a network of liaison services to assist the family as the disease progresses.

Prior to the landmark book, ''On Death and Dying,'' by Kübler-Ross,[7] it was assumed that people who are dying want to avoid acknowledging or discussing their terminal condition. Acknowledgment of being terminal was thought to be tantamount to the death of hope. Contemporary wisdom was that ''after hope dies, the patient dies.'' More recently, it has been stated that the terminally ill wish to talk about death and their own dying, and require the emotional support of professionals and family in order to die a peaceful death.[6] The purpose of this study was to gather data from a population with Duchenne muscular dystrophy (DMD) regarding perceived need and availability of psychosocial services related to the dying process.

The Duchenne form of muscular dystrophy is a genetically determined primary degenerative myopathy. It is inherited as an x-linked recessive gene and passed to boys by their relatively unaffected mothers. Up to 33% of the cases appear to be due to spontaneous mutation either in the patient or his mother.[2] Progressive weakness usually begins in the second year of life when the patient experiences difficulty standing and walking. Most patients become wheelchair bound by age 10. Death usually occurs between ages 20 and 30 due to respiratory infection or cardiac failure. Although the physical aspects of pseudohypertrophic muscular dystrophy have received continued attention since its identification in 1868 by Duchenne, the psychologic needs of patients and their families have been studied less intensively.[8,9,11]

In 1970, Schowalter reviewed some approaches to working with families with a terminally ill child.[10] He stated that ''although it is probably true that a majority of dying adolescents sense their fate, it remains a crucial question whether or not they should be told what they already suspect.'' He cited a study by Vernick[12] that concluded that children nine years and over with leukemia benefited from knowing their diagnosis. However, he noted that at least two experienced pediatricians published articles advocating ''never telling'' a fatally ill child the true diagnosis.[4,5] Schowalter concluded that ''such an important decision must be individualized and based on the child's requests, the parents' consent, and the staff's ability to support either decision.''

In 1979, Buchanan and associates[3] interviewed parents of children with DMD and found that 76% of the families identified psychologic rather then physical issues as their major problem. The most prominent of these were anxiety related to unexpected degeneration of the child's function, social stigma, dread of having to explain DMD to their sons, and guilt about the genetic aspects of the disease. A common complaint from parents was that they were inadequately educated by professionals about DMD. Anxiety and fear raised by misinformation were a burdensome addition to the existing psychologic trauma for parents. Twenty-four percent of patients had not been told they had DMD; others had partial information. It seemed that the children who were without adequate information about their disease demonstrated greater emotion and behavior difficulties. Recommendations from the study included providing more parent education about DMD, increased information to the child about his illness, and greater awareness of the responses of other family members.

The present study was part of a larger investigation of the attitudes of patients with neuromuscular disorders, their families, and professionals who work with them. Specific items related to issues of death and dying were embedded in a more general needs assessment questionnaire. Responses were compared between the three respondent groups, and subpopulations were identified by the child's age for further comparisons.

The study was primarily designed as an exploratory survey of expressed needs. A review of the literature, clinical experience, and anecdotal evidence suggested several interesting questions:

1. How important is information and counseling regarding the dying process to patients and parents? Do these groups perceive such services as available to them?
2. Do professionals perceive these services as more available than patients and parents indicate?
3. Is rated importance of these services age–related? That is, do older patients and their parents value life expectancy information and counseling more highly than younger patients and their parents?
4. Is there congruence between the attitudes of individual patients and their parents on the importance of these services?
5. Would parents of younger children see themselves as the

This study was supported in part by the Muscular Dystrophy Association of America.

Submitted for publication September 9, 1982. Accepted in revised form March 16, 1983.

appropriate target for information and counseling? Would parents of older patients want the entire family to receive these services?

METHOD

Participants

Participants included 32 patients with DMD, 75 parents of children with DMD, and 43 professionals. Responses to a needs assessment questionnaire were gathered during onsite visits to five Muscular Dystrophy Association of America (MDAA) clinics during a six-month period from September, 1980 to March, 1981. Patients 12 years of age and older were asked to participate with parental consent. Patients ranged in age from 12 to 21 years with a mean age of 16.8 years. All but two were wheelchair–bound. Five patients were not interviewed due to overt mental retardation.

Forty percent (17) of the professionals were physicians. Of the remaining personnel, 26% (12) were allied health professionals (nurses, physical therapists, occupational therapists, orthotists); 17% (7) were mental health personnel (psychiatrists, psychologists, social workers); and 17% (7) were administrators (clinic coordinators, MDAA representatives). Experience with muscular dystrophy patients varied from one month to ten years with a median of three and a half years.

Instrument and Procedure

Participants were asked to complete a survey questionnaire regarding the importance and availability of 32 medical and psychosocial services. Patients and parents rated the importance of each service on a 5 point Likert-type scale (1 = very unimportant to 5 = very important). All groups were asked to indicate the degree they perceived availability of services. Patients and parents were asked to designate whether each service had ever been offered to them by the clinic, while professionals were asked which services their clinic was presently providing. This study focused on those responses dealing with psychosocial adaptation to the disease and the dying process.

RESULTS

Service Importance and Availability

Ranked importance and perceived availability of life expectancy information and counseling services for patient, parent, and clinic personnel groups were determined.

A majority of parents (55%) and patients (53%) indicated that life expectancy information was important. Both groups also stated that information about the progress of the disease throughout the life cycle was important. Parents ranked both short- and long-term counseling as more important than did patients. Sixty-five percent of the parents felt short-term counseling (three sessions or less) was an important need, while 45% of parents also viewed long-term counseling (more than three sessions) as important. A relatively small percentage of patients (17% and 13%, respectively) indicated these counseling services were important.

Only a minority of parents and patients perceived either type of information as being available. However, 74% of clinic personnel indicated that data regarding life expectancy were available through their clinics, while nearly half (49%) also reported that life cycle information was available. Neither parents nor patients viewed either type of counseling as generally available. Twenty-two percent of the patients and 17% of parents reported that they had received or been offered either short- or long-term counseling. Seventy-seven percent of clinic personnel indicated that short-term counseling was available through their clinics, while 33% indicated that long-term counseling was also provided.

Congruence between patient and parent attitudes regarding service importance was evaluated by calculations of coefficient correlations. No significant correspondence was found between patient and parent attitudes regarding life span and dying information or short- and long-term counseling. However, patients and their parents demonstrated significant agreement regarding the importance of long-term counseling [Pearson's $r(32) = 0.398, p = 0.05$]. Congruence between patient and parent also approached significance on the importance of life expectancy information [Pearson's $r(34) = 0.313, p = 0.10$]. There were no consistent differences between patient-parent sets which were congruent and those which were not when demographic factors, such as patient or parent age and education level, sex or marital status, were considered.

Differences Related to Patient Age

In order to compare age-related responses, parent participants were subdivided into two groups, those having children 12 years and older (34) and those having children under the age of 12 (41). Both groups were highly similar in demographic characteristics. A comparison between the two parent groups on ranked importance of life expectancy information and counseling services indicated no significant differences. Both groups perceived information and counseling as important.

In order to compare attitudes of younger and older patients, a coefficient correlation (Pearson's r) was run on patient age, ranked importance of life expectancy, life span, and dying information. There was a positive relationship in each case (0.12 and 0.23, respectively), but neither correlation reached significance. Correlations on age and ranked importance of short- and long-term counseling indicated nonsignificant negative relationships (-0.17 and -0.13, respectively).

Thus, it appears that whether a patient or his family want to deal with life and death issues by obtaining information or counseling is an individual matter unrelated to patient age.

In addition to ranking importance of services, the two parent groups were also asked to indicate whom they regarded as the most appropriate primary target for services: themselves, the patient, or the entire family. Primary targets for information and counseling services indicated by the two parent groups is presented in the table. No differences existed between parents of younger and older children with regard to primary targets for life expectancy information. Both parent groups indicated that such information should be given to themselves or to the entire family. However, chi square analysis revealed a significant difference between parent groups in designated primary targets for counseling services ($\chi^2 = 11.62, p = 0.01$). Parents of younger children more frequently designated themselves or the patient as a primary target for counseling services, while

Primary Service Targets Indicated by Parents of Younger Patients (41) and Parents of Older Patients (34)

	Self		Patient		Entire family	
	n	%	*n*	%	*n*	%
Life expectancy and life cycle information						
Parents of younger children	14	34	2	5	20	49
Parents of older children	14	41	3	9	14	41
Counseling services						
Parents of younger children	15	37	14	34	12	29
Parents of older children	3	9	10	29	21	62

parents of older children more frequently designated the entire family to be in need of counseling services. Thus, it appears that as the patient becomes increasingly incapacitated the stress may impact not only on the patient and his parents but on the entire family system.

DISCUSSION

These results point to three major conclusions. First, most patients and parents want education and counseling while dealing with the physical decline from DMD. Second, a significant minority of persons indicates that this is not important to them. Third, professionals perceive these services as being more available than do patients and parents.

Parents have concerns about how to relate to the patient and other family members and friends. They need information on how to locate schools and special programs, how to find and finance suitable medical care, and how to ensure appropriate peer relationships for the patient and provide a balanced family life. Additionally, parents may need help in considering the risk of having another child. Siblings may need help with feelings of jealousy because the affected child seems to get more attention. And as the disabled child matures, he will need information about his disease, his progressive disability, his prognosis, as well as assistance to develop his capacities and potential to the fullest degree in physical, academic, social, emotional, and vocational areas.

There may be uncertainty as to the appropriateness of intervention, whether parents should talk about diagnosis and prognosis with their child, or whether there is a role for professionals; whether it is best to wait for patients and families to bring up such topics or whether professionals should initiate discussions.

Care of the dying patient with dystrophy does not revolve around telling or not telling him that he is dying. He knows he is deteriorating. He receives information from multiple sources, including the physician, nurses, therapists, social workers, family, friends, and clergy, and he observes changes in the behavior of others toward him. He also learns from television, telethons, books and newspapers, and by observing other patients with dystrophy in school or the MDAA clinic. This is not to say that there should be a brutal attack on denial of death in the name of openness. The right to know is counterbalanced by the right not to know. It is important to honor a patient's denial for as long as it is functional for him. However, through being emotionally accessible to the patient over time, the physician and the team can be available when the patient chooses to acknowledge the possibility of death and his fear of dying.

When addressing patient and family needs, it is helpful to recognize the needs of the medical staff also. To talk with, work with, and understand the dying person evokes intense personal feelings. Annas[1] noted that 90% of all cancer patients prefer to know their diagnosis, even if terminal, whereas 60% to 90% of their physicians oppose telling them. The three most frequently observed physician reactions to the death of a patient are anger, denial, and depression. These are precisely the three most common reactions of patients to their learning of death.

Since professionals can experience patient death as an unacceptable conclusion to the health-care process, they may avoid discussing emotionally loaded issues, even though services related to the dying process are theoretically available. Hence, a dying person may receive less personal attention than attention to joint range of motion, grades of muscle tests, or repair of braces.

Several alternative explanations can be postulated for the differing perceptions of service availability between consumers and professionals. The question originally posed to the two groups was not identical. Patients and parents were asked whether such services had ever been offered, while clinic personnel were asked if their clinic presently provided the service. Given this variation, some differences in perception were anticipated. However, actual differences obtained were so consistently large across the service categories that this factor alone would not account for the variation.

Information regarding life expectancy and progressive disability is generally provided to parents at the time of diagnosis or shortly thereafter. Weisman[13] suggested that there may be three clinical phases of living-dying: 1) the acute crisis phase when diagnosis is first made; 2) the chronic living-dying phase; and 3) the terminal phase. At the time of initial diagnosis it is appropriate to deal with the crisis of realizing decreased life expectancy and progressive disability. Yet, simply quoting figures to parents does not adequately convey information to them. Given the emotional trauma of the experience, many parents may not hear what has been said. They may not remember the facts or attempt to deny them. Multiple discussions over time can expand and clarify realities, elicit parents' responses, and help them share their feelings of guilt, denial, depression, and withdrawal. Crucial concerns, such as prognosis—what is going to happen as the child develops—also need to be addressed so that the patient and family can move into a lifestyle and life expectation that takes into account the disease process.

CONCLUSION

The foregoing discussion suggests that several changes might need to occur in the current service delivery system. Time and staffing patterns of MDAA clinics might be restructured to increase availability of psychosocial services to patients and their families. This focusing of psychosocial service delivery appears to be particularly important for the entire family as the patient grows older and becomes more debilitated. Liaisons might also be encouraged between clinics and other resources, such as schools and mental health centers, to provide a network of supportive services within the community. Such a network may afford families more options in dealing with the many adaptive issues they face in adjusting to the demands of the disease process.

Acknowledgment: The authors wish to express their appreciation for the support of David S. Campion, MD, University of California, Los Angeles.

ADDRESS REPRINT REQUESTS TO:
Julie G. Botvin Madorsky, MD
Casa Colina Hospital for Rehabilitative Medicine
255 East Bonita Avenue
Pomona, CA 91767

References

1. Annas GJ: Rights of terminally ill patient. J Nurs Adm **4:**40–44, Mar–Apr 1974
2. Brooke MH: Clinician's View of Neuromuscular Diseases. Baltimore, Williams & Wilkins Co, 1977
3. Buchanan DC, LaBarbera CJ, Roelofs R, Olson W: Reactions of families to children with Duchenne muscular dystrophy. Gen Hosp Psychiatry **1**:262–268, Sept 1979.
4. Evans AE: If a child must die… N Engl J Med **278:**138–142, 1968
5. Howell DA: A child dies. J Pediatr Surg **1:**2–7, 1966
6. Kalish RA: A little myth is a dangerous thing: Research in the service of the dying. *In* Garfield CA (ed): Psychosocial Care of the Dying Patient. San Francisco, University of California School of Medicine, 1976
7. Kübler-Ross E: On Death and Dying. New York, MacMillan, 1969
8. Morrow RS, Cohen J: Psychosocial factors in muscular dystrophy. Child Psychiatry **3:**70–80, 1954
9. Schoelly ML, Fraser AW: Emotional reactions in muscular dystrophy. Am J Phys Med **34:**119–123, 1955
10. Schowalter JE: Child's reaction to his own terminal illness. *In* Schoenberg B, Carr AC, Peretz D, Kutscher AH (eds): Loss and Grief: Psychological Management in Medical Practice. NY, Columbia Press, 1970, pp 51–69
11. Taft LT: Care and management of child with muscular dystrophy. Dev Med Child Neurol **15:**510–518, 1973
12. Vernick J, Karon M: Who's afraid of death on leukemia ward? Am J Dis Child **109:**393–397, 1965
13. Weisman AD: Psychosocial consideration in terminal care. *In* Schoenberg B, Carr AC, Peretz D, Kutscher AH (eds): Psychosocial Aspects of Terminal Care. New York, Columbia University Press, 1972, pp 162–172

EXHIBIT 9

HIGHLIGHTS OF PRESCRIBING INFORMATION

These highlights do not include all the information needed to use EXONDYS 51 safely and effectively. See full prescribing information for EXONDYS 51.

EXONDYS 51 (eteplirsen) injection, for intravenous use
Initial U.S. Approval: 2016

RECENT MAJOR CHANGES

Warnings and Precautions , Hypersensitivity Reactions (5.1) 1/2022

INDICATIONS AND USAGE

EXONDYS 51 is an antisense oligonucleotide indicated for the treatment of Duchenne muscular dystrophy (DMD) in patients who have a confirmed mutation of the DMD gene that is amenable to exon 51 skipping. This indication is approved under accelerated approval based on an increase in dystrophin in skeletal muscle observed in some patients treated with EXONDYS 51 *[see Clinical Studies (14)]*. Continued approval for this indication may be contingent upon verification of a clinical benefit in confirmatory trials. (1)

DOSAGE AND ADMINISTRATION

- 30 milligrams per kilogram of body weight once weekly (2.1)
- Administer as an intravenous infusion over 35 to 60 minutes via an in-line 0.2 micron filter (2.1, 2.3)
- Dilution required prior to administration (2.2)

DOSAGE FORMS AND STRENGTHS

Injection:

- 100 mg/2 mL (50 mg/mL) in single-dose vial (3)
- 500 mg/10 mL (50 mg/mL) in single-dose vial (3)

CONTRAINDICATIONS

None (4)

WARNINGS AND PRECAUTIONS

Hypersensitivity Reactions: Hypersensitivity reactions, including bronchospasm, chest pain, cough, tachycardia, and urticaria, have occurred in patients treated with EXONDYS 51. If hypersensitivity reactions occur, institute appropriate medical treatment and consider slowing the infusion or interrupting the EXONDYS 51 therapy. (2.3, 5.1)

ADVERSE REACTIONS

The most common adverse reactions (incidence ≥35% and higher than placebo) were balance disorder and vomiting. (6.1)

To report SUSPECTED ADVERSE REACTIONS, contact Sarepta Therapeutics, Inc. at 1-888-SAREPTA (1-888-727-3782) or FDA at 1-800-FDA-1088 or *www.fda.gov/medwatch*.

See 17 for PATIENT COUNSELING INFORMATION.

Revised: 1/2022

FULL PRESCRIBING INFORMATION: CONTENTS*

*Sections or subsections omitted from the full prescribing information are not listed.

FULL PRESCRIBING INFORMATION

1 INDICATIONS AND USAGE

EXONDYS 51 is indicated for the treatment of Duchenne muscular dystrophy (DMD) in patients who have a confirmed mutation of the DMD gene that is amenable to exon 51 skipping. This indication is approved under accelerated approval based on an increase in dystrophin in skeletal muscle observed in some patients treated with EXONDYS 51 *[see Clinical Studies (14)]*. Continued approval for this indication may be contingent upon verification of a clinical benefit in confirmatory trials.

2 DOSAGE AND ADMINISTRATION

2.1 Dosing Information

The recommended dose of EXONDYS 51 is 30 milligrams per kilogram administered once weekly as a 35 to 60 minute intravenous infusion via an in-line 0.2 micron filter.

If a dose of EXONDYS 51 is missed, it may be administered as soon as possible after the scheduled time.

2.2 Preparation Instructions

EXONDYS 51 is supplied in single-dose vials as a preservative-free concentrated solution that requires dilution prior to administration. Parenteral drug products should be inspected visually for particulate matter and discoloration prior to administration, whenever solution and container permit. Use aseptic technique.

a. Calculate the total dose of EXONDYS 51 to be administered based on the patient's weight and the recommended dose of 30 milligrams per kilogram. Determine the volume of EXONDYS 51 needed and the correct number of vials to supply the full calculated dose.

b. Allow vials to warm to room temperature. Mix the contents of each vial by gently inverting 2 or 3 times. Do not shake.

c. Visually inspect each vial of EXONDYS 51. EXONDYS 51 is a clear, colorless solution that may have some opalescence, and may contain trace amounts of small, white to off-white amorphous particles. Do not use if the solution in the vials is cloudy, discolored or contains extraneous particulate matter other than trace amounts of small, white to off-white amorphous particles.

d. With a syringe fitted with a 21-gauge or smaller non-coring needle, withdraw the calculated volume of EXONDYS 51 from the appropriate number of vials.

e. Dilute the withdrawn EXONDYS 51 in 0.9% Sodium Chloride Injection, USP, to make a total volume of 100-150 mL. Visually inspect the diluted solution. Do not use if the solution is cloudy, discolored or contains extraneous particulate matter other than trace amounts of small, white to off-white amorphous particles.

f. Administer the diluted solution via an in-line 0.2 micron filter.

g. EXONDYS 51 contains no preservatives and should be administered immediately after dilution. Complete infusion of diluted EXONDYS 51 solution within 4 hours of dilution. If immediate use is not possible, the diluted solution may be stored for up to 24 hours at 2ºC to 8ºC (36ºF to 46ºF). Do not freeze. Discard unused EXONDYS 51.

2.3 Administration Instructions

Application of a topical anesthetic cream to the infusion site prior to administration of EXONDYS 51 may be considered.

EXONDYS 51 is administered via intravenous infusion. Flush the intravenous access line with 0.9% Sodium Chloride Injection, USP, prior to and after infusion.

Infuse the diluted EXONDYS 51 solution over 35 to 60 minutes via an in-line 0.2 micron filter. Do not mix other medications with EXONDYS 51 or infuse other medications concomitantly via the same intravenous access line.

If a hypersensitivity reaction occurs, consider slowing the infusion or interrupting the EXONDYS 51 therapy *[see Warnings and Precautions (5.1) and Adverse Reactions (6.1)]*.

3 DOSAGE FORMS AND STRENGTHS

EXONDYS 51 is a clear and colorless solution that may have some opalescence, and may contain trace amounts of small, white to off-white amorphous particles, and is available as follows:

- Injection: 100 mg/2 mL (50 mg/mL) solution in a single-dose vial
- Injection: 500 mg/10 mL (50 mg/mL) solution in a single-dose vial

4 CONTRAINDICATIONS

None.

5 WARNINGS AND PRECAUTIONS

5.1 Hypersensitivity Reactions

Hypersensitivity reactions, including bronchospasm, chest pain, cough, tachycardia, and urticaria, have occurred in patients who were treated with EXONDYS 51. If a hypersensitivity reaction occurs, institute appropriate medical treatment and consider slowing the infusion or interrupting the EXONDYS 51 therapy *[see Dosage and Administration (2.3)]*.

6 ADVERSE REACTIONS

6.1 Clinical Trials Experience

Because clinical trials are conducted under widely varying conditions, adverse reaction rates observed in clinical trials of a drug cannot be directly compared to rates in the clinical trials of another drug and may not reflect the rates observed in practice.

EXONDYS 51 was studied in a double-blind, placebo-controlled study for 24 weeks (Study 1), followed by an open-label extension (Study 2). In Study 1, 12 patients were randomized to receive weekly intravenous infusions of EXONDYS 51 (n=8) or placebo (n=4) for 24 weeks. All 12 patients continued in Study 2 and received open-label EXONDYS 51 weekly for up to 208 weeks

In Study 1, 4 patients received placebo, 4 patients received EXONDYS 51 30 mg/kg, and 4 patients received EXONDYS 51 50 mg/kg (1.7 times the recommended dosage). In Study 2, 6 patients received EXONDYS 51 30 mg/kg/week and 6 patients received EXONDYS 51 50 mg/kg/week *[see Clinical Studies (14)]*.

Adverse reactions that occurred in 2 or more patients who received EXONDYS 51 and were more frequent than in the placebo group in Study 1 are presented in Table 1 (the 30 and 50 mg/kg groups are pooled). Because of the small numbers of patients, these represent crude frequencies that may not reflect the frequencies observed in practice. The 50 mg/kg once weekly dosing regimen of EXONDYS 51 is not recommended *[see Dosage and Administration (2.1)]*.

The most common adverse reactions were balance disorder and vomiting.

Table 1. Adverse Reactions in DMD Patients Treated with 30 or 50 mg/kg/week[1] EXONDYS 51 with Incidence at Least 25% More than Placebo (Study 1)

Adverse Reactions	EXONDYS 51 (N=8) %	Placebo (N=4) %
Balance disorder	38	0
Vomiting	38	0
Contact dermatitis	25	0

[1] 50 mg/kg/week = 1.7 times the recommended dosage

Adverse Reactions from Observational Clinical Studies

The following adverse reactions have been identified during observational studies that were conducted as part of the clinical development program and continued postapproval.

In open-label observational studies, 163 patients received at least one intravenous dose of EXONDYS 51, with doses ranging between 0.5 mg/kg (0.017 times the recommended dosage) and 50 mg/kg (1.7 times the recommended dosage). All patients were male and had genetically confirmed Duchenne muscular dystrophy. Age at study entry was 6 months to 19 years. Most (85%) patients were Caucasian.

The most common adverse reactions seen in greater than 10% of the study population were headache, cough, rash, and vomiting.

Hypersensitivity reactions have occurred in patients treated with EXONDYS 51 *[see Warnings and Precautions (5.1)]*.

6.2 Postmarketing Experience

The following adverse reactions have been identified during postapproval use of EXONDYS 51. Because these reactions are reported voluntarily from a population of uncertain size, it is not always possible to reliably estimate their frequency or establish a causal relationship to drug exposure.

Postmarketing adverse reactions that occurred during infusion include bronchospasm, cyanosis of the lips, and malaise. The following adverse reactions have also been reported in patients receiving EXONDYS 51: pyrexia, flushing, protein urine present, and dehydration.

8 USE IN SPECIFIC POPULATIONS

8.1 Pregnancy

Risk Summary

There are no human or animal data available to assess the use of EXONDYS 51 during pregnancy. In the U.S. general population, major birth defects occur in 2 to 4% and miscarriage occurs in 15 to 20% of clinically recognized pregnancies.

8.2 Lactation

Risk Summary

There are no human or animal data to assess the effect of EXONDYS 51 on milk production, the presence of eteplirsen in milk, or the effects of EXONDYS 51 on the breastfed infant.

The developmental and health benefits of breastfeeding should be considered along with the mother's clinical need for EXONDYS 51 and any potential adverse effects on the breastfed infant from EXONDYS 51 or from the underlying maternal condition.

8.4 Pediatric Use

EXONDYS 51 is indicated for the treatment of Duchenne muscular dystrophy (DMD) in patients who have a confirmed mutation of the DMD gene that is amenable to exon 51 skipping, including pediatric patients *[see Clinical Studies (14)]*.

Intravenous administration of eteplirsen (0, 100, 300, or 900 mg/kg) to juvenile male rats once weekly for 10 weeks beginning on postnatal day 14 resulted in renal tubular necrosis at the highest dose tested and decreased bone densitometry parameters (mineral density, mineral content, area) at all doses. The kidney findings were associated with clinical pathology changes (increased serum urea nitrogen and creatinine, decreased urine creatinine clearance). No effects were observed on the male reproductive system, neurobehavioral development, or immune function. An overall no-effect dose was not identified. Plasma eteplirsen exposure (AUC) at the

lowest dose tested (100 mg/kg) was similar to that in humans at the recommended human dose (30 mg/kg).

8.5 Geriatric Use

DMD is largely a disease of children and young adults; therefore, there is no geriatric experience with EXONDYS 51.

8.6 Patients with Renal Impairment

Renal clearance of eteplirsen is reduced in non-DMD adults with renal impairment based on estimated creatinine clearance *[see Clinical Pharmacology (12.3)]*. However, because of the effect of reduced skeletal muscle mass on creatinine measurements in DMD patients, no specific dosage adjustment can be recommended for DMD patients with renal impairment.

10 OVERDOSAGE

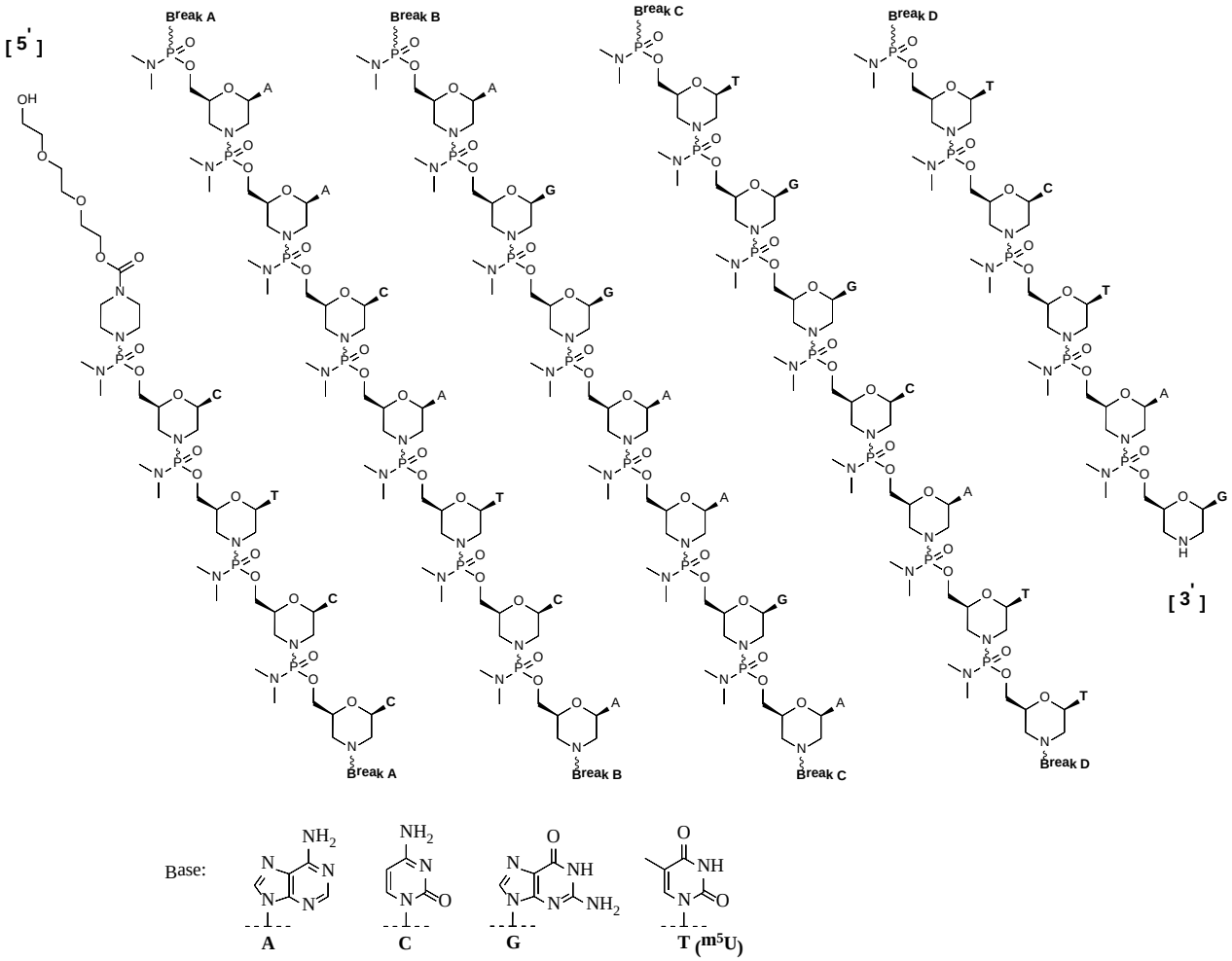
There is no experience with overdose of EXONDYS 51.

11 DESCRIPTION

EXONDYS 51 (eteplirsen) injection is a sterile, aqueous, preservative-free, concentrated solution for dilution prior to intravenous administration. EXONDYS 51 is clear and colorless, and may have some opalescence, and may contain trace amounts of small, white to off-white amorphous particles. EXONDYS 51 is supplied in single dose vials containing 100 mg or 500 mg eteplirsen (50 mg/mL). EXONDYS 51 is formulated as an isotonic, phosphate buffered saline solution with an osmolality of 260 to 320 mOsm and a pH of 7.5. Each milliliter of EXONDYS 51 contains 50 mg eteplirsen; 0.2 mg potassium chloride, 0.2 mg potassium phosphate monobasic, 8 mg sodium chloride, and 1.14 mg sodium phosphate dibasic, anhydrous, in water for injection. The product may contain hydrochloric acid or sodium hydroxide to adjust pH.

Eteplirsen is an antisense oligonucleotide of the phosphorodiamidate morpholino oligomer (PMO) subclass. PMOs are synthetic molecules in which the five-membered ribofuranosyl rings found in natural DNA and RNA are replaced by a six-membered morpholino ring. Each morpholino ring is linked through an uncharged phosphorodiamidate moiety rather than the negatively charged phosphate linkage that is present in natural DNA and RNA. Each phosphorodiamidate morpholino subunit contains one of the heterocyclic bases found in DNA (adenine, cytosine, guanine, or thymine). Eteplirsen contains 30 linked subunits. The molecular formula of eteplirsen is $C_{364}H_{569}N_{177}O_{122}P_{30}$ and the molecular weight is 10305.7 daltons.

The structure and base sequence of eteplirsen are:



The sequence of bases from the 5' end to the 3' end is:
CTCCAACATCAAGGAAGATGGCATTTCTAG

12 CLINICAL PHARMACOLOGY

12.1 Mechanism of Action

Eteplirsen is designed to bind to exon 51 of dystrophin pre-mRNA, resulting in exclusion of this exon during mRNA processing in patients with genetic mutations that are amenable to exon 51 skipping. Exon skipping is intended to allow for production of an internally truncated dystrophin protein, which was evaluated in Study 2 and Study 3 *[see Clinical Studies (14)]*.

12.2 Pharmacodynamics

All EXONDYS 51-treated patients evaluated (n=36) were found to produce messenger ribonucleic acid (mRNA) for a truncated dystrophin protein by reverse transcription polymerase chain reaction.

In Study 2, the average dystrophin protein level in muscle tissue after 180 weeks of treatment with EXONDYS 51 was 0.93% of normal (i.e., 0.93% of the dystrophin level in healthy subjects). Because of insufficient information on dystrophin protein levels before treatment with EXONDYS 51 in Study 1, it is not possible to estimate dystrophin production in response to EXONDYS 51 in Study 1.

In Study 3, the average dystrophin protein level was 0.16% of normal before treatment, and 0.44% of normal after 48 weeks of treatment with EXONDYS 51 *[see Clinical Studies (14)]*. The median increase in truncated dystrophin in Study 3 was 0.1% *[see Clinical Studies (14)]*.

Dystrophin levels assessed by western blot can be meaningfully influenced by differences in sample processing, analytical technique, reference materials, and quantitation methodologies. Therefore, comparing dystrophin results from different assay protocols will require a standardized reference material and additional bridging studies.

12.3 Pharmacokinetics

Following single or multiple intravenous infusions of EXONDYS 51 in male pediatric DMD patients, plasma concentration-time profiles of eteplirsen were generally similar and showed multi-phasic decline. The majority of drug elimination occurred within 24 hours. Approximate dose-proportionality and linearity in PK properties were observed following multiple-dose studies (0.5 mg/kg/week [0.017 times the recommended dosage] to 50 mg/kg/week [1.7 times the recommended dosage]). There was no significant drug accumulation following weekly dosing across this dose range. The inter-subject variability for eteplirsen C_{max} and AUC range from 20 to 55%.

Following single or multiple intravenous infusions of EXONDYS 51, the peak plasma concentrations (C_{max}) of eteplirsen occurred near the end of infusion (i.e., 1.1 to 1.2 hours across a dose range of 0.5 mg/kg/week to 50 mg/kg/week).

Distribution

In vitro investigation suggested that plasma protein binding of eteplirsen in human ranges between 6 to 17%. The mean apparent volume of distribution (Vss) of eteplirsen was 600 mL/kg following weekly intravenous infusion of EXONDYS 51 at 30 mg/kg.

Twenty-four hours after the end of the infusion, mean concentrations of eteplirsen were 0.07% of C_{max}. Accumulation of eteplirsen during once weekly dosing has not been observed.

Elimination

The total clearance of eteplirsen was 339 mL/hr/kg following 12 weeks of therapy with 30 mg/kg/week.

Metabolism

Eteplirsen did not appear to be metabolized by hepatic microsomes of any species tested, including humans.

Excretion

Renal clearance of eteplirsen accounts for approximately two-thirds of the administered dose within 24 hours of intravenous administration. Elimination half-life ($t_{1/2}$) of eteplirsen was 3 to 4 hours.

Specific Populations

Age:

The pharmacokinetics of eteplirsen have been evaluated in male pediatric DMD patients. There is no experience with the use of EXONDYS 51 in patients 65 years of age or older.

Sex:

Sex effects have not been evaluated; EXONDYS 51 has not been studied in female patients.

Race:

Potential impact of race is not known because 89% of the patients in studies were Caucasians.

Patients with Renal Impairment:

The effect of renal impairment on the pharmacokinetics of eteplirsen was evaluated in non-DMD subjects aged 51 to 75 years with mild (n=8, creatinine clearance ≥60 mL/min and <90 mL/min) or moderate (n=8, creatinine clearance ≥30 mL/min and <60 mL/min) renal impairment and matched healthy subjects (n=9, creatinine clearance >90 mL/min). Subjects received a single 30 mg/kg intravenous dose of eteplirsen.

Subjects with mild and moderate renal impairment showed higher eteplirsen exposure compared to subjects with normal renal function. In subjects with mild and moderate renal impairment, exposure (AUC) increased approximately 1.4-fold and 2.4-fold, respectively. The effect of severe renal impairment or end-stage renal disease on eteplirsen pharmacokinetics and safety has not been studied.

Estimated creatinine clearance values derived from the Cockcroft-Gault equation and the threshold definitions for mild, moderate, and severe renal impairment in otherwise healthy adults would not be generalizable to patients with DMD. Therefore, no specific dosage adjustment can be recommended for patients with renal impairment.

Patients with Hepatic Impairment:

EXONDYS 51 has not been studied in patients with hepatic impairment.

Drug Interaction Studies

In vitro data showed that eteplirsen did not significantly inhibit CYP1A2, CYP2B6, CYP2C8, CYP2C9, CYP2C19, CYP2D6, or CYP3A4/5. Eteplirsen did not induce CYP2B6 or CYP3A4, and induction of CYP1A2 was substantially less than the prototypical inducer, omeprazole. Eteplirsen was not a substrate nor did it have any major inhibitory potential for any of the key human transporters tested (OAT1, OAT3, OCT1, OCT2, OATP1B1, OATP1B3, P-gp, BCRP, MRP2 and BSEP). Based on *in vitro* data on plasma protein binding, CYP or drug transporter interactions, and microsomal metabolism, eteplirsen is expected to have a low potential for drug-drug interactions in humans.

13 NONCLINICAL TOXICOLOGY

13.1 Carcinogenesis, Mutagenesis, Impairment of Fertility

Carcinogenesis

Administration of eteplirsen to male transgenic (Tg.rasH2) mice (0, 200, 500, or 960 mg/kg) weekly for 26 weeks (intravenous [IV] injection for 15 weeks, followed by subcutaneous injection for 11 weeks) and to male rats (0, 60, 180, or 600 mg/kg IV) weekly for 96 weeks resulted in no increase in neoplasms.

Mutagenesis

Eteplirsen was negative in *in vitro* (bacterial reverse mutation and chromosomal aberration in CHO cells) and *in vivo* (mouse bone marrow micronucleus) assays.

Impairment of Fertility

Fertility studies in animals were not conducted with eteplirsen. No effects on the male reproductive system were observed following intravenous administration of eteplirsen (0, 5, 40, or 320 mg/kg) to male monkeys once weekly for 39 weeks. Plasma eteplirsen exposure (AUC) in monkeys at the highest dose tested was 20 times that in humans at recommended human dose (30 mg/kg).

14 CLINICAL STUDIES

EXONDYS 51 was evaluated in three clinical studies in patients who have a confirmed mutation of the DMD gene that is amenable to exon 51 skipping.

In Study 1, patients were randomized to receive weekly infusions of EXONDYS 51 (30 mg/kg, n=4); EXONDYS 51 (50 mg/kg, n=4), or placebo (n=4) for 24 weeks. The primary endpoint was dystrophin production; a clinical outcome measure, the 6-minute walk test (6MWT), was also assessed. The 6MWT measures the distance that a patient can walk on a flat, hard surface in a period of 6 minutes. Patients had a mean age of 9.4 years, a mean 6-minute walk distance (6MWD) at baseline of 363 meters, and were on a stable dose of corticosteroids for at least 6 months. There was no significant difference in change in 6MWD between patients treated with EXONDYS 51 and those treated with placebo.

All 12 patients who participated in Study 1 continued treatment with open-label EXONDYS 51 weekly for an additional 4 years in Study 2. The 4 patients who had been randomized to placebo were re-randomized 1:1 to EXONDYS 51 30 or 50 mg/kg/week such that there were 6 patients on each dose. Patients who participated in Study 2 were compared to an external control group. The primary clinical efficacy outcome measure was the 6MWT. Eleven patients in Study 2 had a muscle biopsy after 180 weeks of treatment with EXONDYS 51, which was analyzed for dystrophin protein level by Sarepta western blot. Study 2 failed to provide evidence of a clinical benefit of EXONDYS 51 compared to the external control group. The average dystrophin protein level after 180 weeks of treatment with EXONDYS 51 was 0.93% of the dystrophin level in healthy subjects. Because of insufficient information on dystrophin protein levels before treatment with EXONDYS 51 in Study 1, it is not possible to estimate dystrophin production in response to EXONDYS 51 in Study 1.

In Study 3, 13 patients were treated with open-label EXONDYS 51 (30 mg/kg) weekly for 48 weeks and had a muscle biopsy at baseline and after 48 weeks of treatment. Patients had a mean age of 8.9 years and were on a stable dose of corticosteroids for at least 6 months. Dystrophin levels in muscle tissue were assessed by Western blot. In the 12 patients with evaluable results, the pre-treatment dystrophin level was 0.16% ± 0.12% (mean ± standard deviation) of the dystrophin level in a healthy subject and 0.44% ± 0.43% after 48 weeks of treatment with EXONDYS 51 ($p < 0.05$). The median increase after 48 weeks was 0.1%.

Individual patient dystrophin levels from Study 3 are shown in Table 2.

Table 2. Sarepta Western Blot Results: EXONDYS 51-Treated (Week 48) vs Pre-treatment Baseline (% Normal Dystrophin) (Study 301)

Patient Number	Baseline % normal dystrophin	Week 48 % normal dystrophin	Change from Baseline % normal dystrophin
1	0.13	0.26	0.13
2	0.35	0.36	0.01
3	0.06	0.37	0.31
4	0.04	0.10	0.06
5	0.17	1.02	0.85
6	0.37	0.30	-0.07
7	0.17	0.42	0.25
8	0.24	1.57	1.33
9	0.11	0.12	0.01
10	0.05	0.47	0.43
11	0.02	0.09	0.07
12	0.18	0.21	0.03
Mean	0.16	0.44	0.28; p=0.008

16 HOW SUPPLIED/STORAGE AND HANDLING

16.1 How Supplied

EXONDYS 51 injection is supplied in single-dose vials. The solution is clear and colorless, and may have some opalescence, and may contain trace amounts of small, white to off-white amorphous particles.

- Single-dose vials containing 100 mg/2 mL (50 mg/mL) eteplirsen NDC 60923-363-02
- Single-dose vials containing 500 mg/10 mL (50 mg/mL) eteplirsen NDC 60923-284-10

16.2 Storage and Handling

Store EXONDYS 51 at 2°C to 8°C (36°F to 46°F). Do not freeze. Protect from light and store EXONDYS 51 in the original carton until ready for use.

17 PATIENT COUNSELING INFORMATION

Hypersensitivity Reactions

Advise patients and/or caregivers that symptoms of hypersensitivity, including bronchospasm, chest pain, cough, tachycardia, and urticaria can occur with EXONDYS 51. Instruct them to seek immediate medical care should they experience signs and symptoms of hypersensitivity *[see Warnings and Precautions (5.1)]*.

Manufactured for:
Sarepta Therapeutics, Inc.
Cambridge, MA 02142 USA

EXHIBIT 10

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 2 206 781 B1

(12) **EUROPEAN PATENT SPECIFICATION**

(45) Date of publication and mention of the grant of the patent:
02.12.2015 Bulletin 2015/49

(21) Application number: **10004274.6**

(22) Date of filing: **28.06.2005**

(51) Int Cl.:
C12N 15/113 *(2010.01)* ***A61K 48/00*** *(2006.01)*
A61K 31/7088 *(2006.01)* ***A61K 31/712*** *(2006.01)*
C07H 21/00 *(2006.01)*

(54) **Antisense oligonucleotides for inducing exon skipping and methods of use thereof**

Antisense-Oligonukleotide zur Induktion von Exon-Skipping sowie Verfahren zur Verwendung davon

Oligonucléotides antisens permettant d'induire un saut d'exon et leurs procédés d'utilisation

(84) Designated Contracting States:
AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HU IE IS IT LI LT LU MC NL PL PT RO SE SI SK TR
Designated Extension States:
HR LV

(30) Priority: **28.06.2004 AU 2004903474**

(43) Date of publication of application:
14.07.2010 Bulletin 2010/28

(60) Divisional application:
12162995.0 / 2 500 430
15168694.6 / 2 933 332
15190341.6

(62) Document number(s) of the earlier application(s) in accordance with Art. 76 EPC:
05754344.9 / 1 766 010

(73) Proprietor: **THE UNIVERSITY OF WESTERN AUSTRALIA**
Crawley, Western Australia 6009 (AU)

(72) Inventors:
- **Wilton, Stephen Donald**
  Applecross,
  Western Australia 6153 (AU)
- **Fletcher, Sue**
  Bayswater,
  Western Australia 6053 (AU)
- **McClorey, Graham**
  Bayswater,
  Western Australia 6053 (AU)

(74) Representative: **Brasnett, Adrian Hugh et al**
Mewburn Ellis LLP
City Tower
40 Basinghall Street
London EC2V 5DE (GB)

(56) References cited:
WO-A1-93/20227 **WO-A2-03/053341**
WO-A2-2004/083446 **CA-A1- 2 507 125**

- **AARTSMA-RUS, A. ET AL.: "Therapeutic antisense-induced exon skipping in cultured muscle cells from six different DMD patients", HUMAN MOLECULAR GENETICS, vol. 12, no. 8, 2003, pages 907-914, XP008084159, ISSN: 0964-6906**
- **AARTSMA-RUS A ET AL: "Targeted exon skipping as a potential gene correction therapy for Duchenne muscular dystrophy", NEUROMUSCULAR DISORDERS,, vol. 12, 1 January 2002 (2002-01-01), pages 71-77, XP008116183, DOI: 10.1016/S0960-8966(02) 00086-X**
- **MATSUO M: "DUCHENNE AND BECKER MUSCULAR DYSTROPHY: FROM GENE DIAGNOSIS TO MOLECULAR THERAPY", IUBMB LIFE, vol. 53, no. 3, 1 March 2002 (2002-03-01), pages 147-152, XP009021242, ISSN: 1521-6543**
- **ERRINGTON STEPHEN J ET AL: "Target selection for antisense oligonucleotide induced exon skipping in the dystrophin gene.", THE JOURNAL OF GENE MEDICINE, vol. 5, no. 6, June 2003 (2003-06), pages 518-527, XP002559309, ISSN: 1099-498X**

Printed by Jouve, 75001 PARIS (FR)

(Cont. next page)

Remarks:

•This application was filed on 22-04-2010 as a divisional application to the application mentioned under INID code 62.

•The file contains technical information submitted after the application was filed and not included in this specification

Description

Field of the Invention

[0001] The present invention relates to novel antisense compounds and compositions suitable for facilitating exon skipping. It also provides methods for inducing exon skipping using the novel antisense compounds as well as therapeutic compositions adapted for use in the methods of the invention.

Background Art

[0002] Significant effort is currently being expended researching methods for suppressing or compensating for disease-causing mutations in genes. Antisense technologies are being developed using a range of chemistries to affect gene expression at a variety of different levels (transcription, splicing, stability, translation). Much of that research has focused on the use of antisense compounds to correct or compensate for abnormal or disease-associated genes in a myriad of different conditions.

[0003] Antisense molecules are able to inhibit gene expression with exquisite specificity and because of this many research efforts concerting oligonucleotides as modulators of gene expression have focused on inhibiting the expression of targeted genes such as oncogenes or viral genes. The antisense oligonucleotides are directed either against RNA (sense strand) or against DNA where they form triplex structures inhibiting transcription by RNA polymerase II. To achieve a desired effect in specific gene down-regulation, the oligonucleotides must either promote the decay of the targeted mRNA or block translation of that mRNA, thereby effectively preventing *de novo* synthesis of the undesirable target protein.

[0004] Such techniques are not useful where the object is to up-regulate production of the native protein or compensate for mutations which induce premature termination of translation such as nonsense or frame-shifting mutations. Furthermore, in cases where a normally functional protein is prematurely terminated because of mutations therein, a means for restoring some functional protein production through antisense technology has been shown to be possible through intervention during the splicing processes (Sierakowska H, et al., (1996) Proc Natl Acad Sci USA 93, 12840-12844; Wilton SD, et al., (1999) Neuromusc Disorders 9, 330-338; van Deutekom JC et al., (2001) Human Mol Genet 10, 1547-1554). In these cases, the defective gene transcript should not be subjected to targeted degradation so the antisense oligonucleotide chemistry should not promote target mRNA decay.

[0005] In a variety of genetic diseases, the effects of mutations on the eventual expression of a gene can be modulated through a process of targeted exon skipping during the splicing process. The splicing process is directed by complex multi-particle machinery that brings adjacent exon-intron junctions in pre-mRNA into close proximity and performs cleavage of phosphodiester bonds at the ends of the introns with their subsequent reformation between exons that are to be spliced together. This complex and highly precise process is mediated by sequence motifs in the pre-mRNA that are relatively short semi-conserved RNA segments to which bind the various nuclear splicing factors that are then involved in the splicing reactions. By changing the way the splicing machinery reads or recognises the motifs involved in pre-mRNA processing, it is possible to create differentially spliced mRNA molecules. It has now been recognised that the majority of human genes are alternatively spliced during normal gene expression, although the mechanism invoked have not been identified. Using antisense oligonucleotides, it has been shown that errors and deficiencies in a coded mRNA could be bypassed or removed from the mature gene transcripts.

[0006] In nature, the extent of genetic deletion or exon skipping in the splicing process is not fully understood, although many instances have been documented to occur, generally at very low levels (Sherrat TG, et al., (1993) Am J Hum Genet 53, 1007-1015). However, it is recognised that if exons associated with disease-causing mutations can be specifically deleted from some genes, a shortened protein product can sometimes be produced that has similar biological properties of the native protein or has sufficient biological activity to ameliorate the disease caused by mutations associated with the target exon (Lu QL, et al., (2003) Nature Medicine 9, 1009-1014; Aartsma-Rus A et al., (2004) Am J Hum Genet 74: 83-92).

[0007] This process of targeted exon skipping is likely to be particularly useful in long genes where there are many exons and introns, where there is redundancy in the genetic constitution of the exons or where a protein is able to function without one or more particular exons (e.g. with the dystrophin gene, which consists of 79 exons; or possibly some collagen genes which encode for repeated blocks of sequence or the huge nebulin or titin genes which are comprised of ~80 and over 370 exons, respectively).

[0008] Efforts to redirect gene processing for the treatment of genetic diseases associated with truncations caused by mutations in various genes have focused on the use of antisense oligonucleotides that either (1) fully or partially overlap with the elements involved in the splicing process; or (2) bind to the pre-mRNA at a position sufficiently close to the element to disrupt the binding and function of the splicing factors that would normally mediate a particular splicing reaction which occurs at that element (e.g., binds to the pre-mRNA at a position within 3, 6, or 9 nucleotides of the

element to be blocked).

[0009] For example, modulation of mutant dystrophin pre-mRNA splicing with antisense oligoribonucleotides has been reported both *in vitro* and *in vivo*. In one type of dystrophin mutation reported in Japan, a 52-base pair deletion mutation causes exon 19 to be removed with the flanking introns during the splicing process (Matsuo et al., (1991) J Clin Invest. 87:2127-2131). An *in vitro* minigene splicing system has been used to show that a 31-mer 2'-O-methyl oligoribonucleotide complementary to the 5' half of the deleted sequence in dystrophin Kobe exon 19 inhibited splicing of wild-type *pre-mRNA* (Takeshima et al. (1995), J. Clin. Invest., 95, 515-520). The same oligonucleotide was used to induce exon skipping from the native dystrophin gene transcript in human cultured lymphoblastoid cells.

[0010] Dunckley et al., (1997) Nucleosides & Nucleotides, 16, 1665-1668 described *in vitro* constructs for analysis of splicing around exon 23 of mutated dystrophin in the *mdx* mouse mutant, a model for muscular dystrophy. Plans to analyse these constructs *in vitro* using 2' modified oligonucleotides targeted to splice sites within and adjacent to mouse dystrophin exon 23 were discussed, though no target sites or sequences were given.

[0011] 2'-O-methyl oligoribonucleotides were subsequently reported to correct dystrophin deficiency in myoblasts from the *mdx* mouse from this group. An antisense oligonucleotide targeted to the 3' splice site of murine dystrophin intron 22 was reported to cause skipping of the mutant exon as well as several flanking exons and created a novel in-frame dystrophin transcript with a novel internal deletion. This mutated dystrophin was expressed in 1-2% of antisense treated mdx myotubes. Use of other oligonucleotide modifications such as 2'-O-methoxyethyl phosphodiesters are described (Dunckley et al. (1998) Human Mol. Genetics, 5, 1083-90).

[0012] Thus, antisense molecules may provide a tool in the treatment of genetic disorders such as Duchenne Muscular Dystrophy (DMD). However, attempts to induce exon skipping using antisense molecules have had mixed success. Studies on dystrophin exon 19, where successful skipping of that exon from the dystrophin pre-mRNA was achieved using a variety of antisense molecules directed at the flanking splice sites or motifs within the exon involved in exon definition as described by Errington et al. (2003) J Gen Med 5, 518-527".

[0013] In contrast to the apparent ease of exon 19 skipping, the first report of exon 23 skipping in the *mdx* mouse by Dunckley *et al*., (1998) is now considered to be reporting only a naturally occurring revertant transcript or artefact rather than any true antisense activity. In addition to not consistently generating transcripts missing exon 23, Durickley *et al*., (1998) did not show any time course of induced exon skipping, or even titration of antisense oligonucleotides, to demonstrate dose dependent effects where the levels of exon skipping corresponded with increasing or decreasing amounts of antisense oligonucleotide. Furthermore, this work could not be replicated by other researchers.

[0014] The first example of specific and reproducible exon skipping in the *mdx* mouse model was reported by Wilton et al., (1999) Neuromuscular Disorders 9, 330-338. By directing an antisense molecule to the donor splice site, consistent and efficient exon 23 skipping was induced in the dystrophin mRNA within 6 hours of treatment of the cultured cells. Wilton *et al*., (1999), also describe targeting the acceptor-region of the mouse dystrophin pre-mRNA with longer antisense oligonucleotides and being unable to repeat the published results of Dunckley *et al*., (1998). No exon skipping, either 23 alone or multiple removal of several flanking exons, could be reproducibly detected using a selection of antisense oligonucleotides directed at the acceptor splice site of intron 22.

[0015] While the first antisense oligonucleotide directed at the intron 23 donor splice site induced consistent exon skipping in primary cultured myoblasts, this compound was found to be much less efficient in immortalized cell cultures expressing higher levels of dystrophin. However, with refined targeting and antisense oligonucleotide design, the efficiency of specific exon removal was increased by almost an order of magnitude (see Mann CJ et al., (2002) J Gen Med 4, 644-654).

[0016] Other disclosures relating to DMD therapy include CA 2507125, Aartsma-Rus et al., Human Molecular Genetics 12, (2003) 907-14, Aartsma-rus et al., Neuromuscular Disorders, 12 (2002) 71-7, as well as WO 2004/083446, published on 30 September 2004.

[0017] Thus, there remains a need to provide antisense oligonucleotides capable of binding to and modifying the splicing of a target nucleotide sequence. Simply directing the antisense oligonucleotides to motifs presumed to be crucial for splicing is no guarantee of the efficacy of that compound in a therapeutic setting.

Summary of the Invention

[0018] The present invention provides an isolated antisense oligonucleotide that binds to human dystrophin pre-mRNA, wherein said oligonucleotide is 20 to 31 nucleotides in length and is an oligonucleotide that is specifically hybridizable to an exon 53 target region of the Dystrophin gene designated as annealing site H53A (+23+47), annealing site H53A (+39+69), or both, wherein said antisense oligonucleotide is a morpholino antisense oligonucleotide, and, wherein said oligonucleotide induces exon 53 skipping..

[0019] The invention further provides a composition comprising an antisense oligonucleotide according to the invention and a saline solution that includes a phosphate buffer.

[0020] The invention further provides an antisense oligonucleotide according to the invention, or a composition ac-

cording to the invention, for use in a method of treatment of muscular dystrophy.

[0021] The invention is further defined in the accompanying claims.

[0022] The choice of target selection plays a crucial role in the efficiency of exon skipping and hence its subsequent application of a potential therapy. Simply designing antisense molecules to target regions of pre-mRNA presumed to be involved in splicing is no guarantee of inducing efficient and specific exon skipping. The most obvious or readily defined targets for splicing intervention are the donor and acceptor splice sites although there are less defined or conserved motifs including exonic splicing enhancers, silencing elements and branch points.

[0023] The acceptor and donor splice sites have consensus sequences of about 16 and 8 bases respectively (see Figure 1 for schematic representation of motifs and domains involved in exon recognition, intron removal and the splicing process).

[0024] The invention also addresses the use of purified and isolated antisense oligonucleotides of the invention, for the manufacture of a medicament for treatment of a genetic disease.

[0025] The invention may be used for treating a condition characterised by Duchenne muscular dystrophy, by administering to a patient in need of treatment an effective amount of an appropriately designed antisense oligonucleotide of the Invention, relevant to the particular genetic lesion in that patient.

[0026] Further, the invention may be used for prophylactically treating a patient to prevent or at least minimize Duchenne muscular dystrophy, comprising the step of: administering to the patient an effective amount of an antisense oligonucleotide or a pharmaceutical composition comprising one or more of these biological molecules.

[0027] Also described herein are kits for treating a genetic disease, which kits comprise at least a antisense oligonucleotide of the present invention, packaged in a suitable container and instructions for its use.

[0028] Other aspects and advantages of the invention will become apparent to those skilled in the art from a review of the ensuing description, which proceeds with reference to the following figures.

Brief Description of the Drawings

[0029]

Figure 1 Schematic representation of motifs and domains Involved in exon recognition, intron removal and the splicing process.

Figure 2 Diagrammatic representation of the concept of antisense oligonucleotide induced exon skipping to by-pass disease-causing mutations (not drawn to scale). The hatched box represents an exon carrying a mutation that prevents the translation of the rest of the mRNA into a protein. The solid black bar represents an antisense oligonucleotide that prevents inclusion of that exon in the mature mRNA.

Figure 3 Gel electrophoresis showing exon 51, exon 52 and exon 53 skipping using various antisense molecules directed at exons 51, 52 and 53, respectively. A "cocktail" of antisense molecules is also shown directed at exon 53

Brief Description of the Sequence Listings

[0030]

Table 1: Description of 2'-O-methyl phosphorothioate antisense oligonucleotides that have been used to date to study induced exon skipping during the processing of the dystrophie pre-mRNA. Since these 2'-O-methyl antisense oligonucleotides are more RNA-like, U represents uracil. With other antisense chemistries such as peptide nucleic acids or morpholinos, these U bases may be shown as "**T**".

SEQ ID	SEQUENCE	NUCLEOTIDE SEQUENCE (5'-3')
192	H53A(+39+62)	CUG UUG CCU CCG GUU CUG AAG GUG
193	H53A(+39+69)	CAU UCA ACU GUU GCC UCC GGU UCU GAA GGU G
195	H53A(+23+47)	CUG AAG GUG UUC UUG UAC UUC AUC C

Detailed Description of the Invention

General

[0031] Those skilled in the art will appreciate that the invention described herein is susceptible to variations and modifications other than those specifically described. It is to be understood that the invention includes all such variation and modifications. The invention also includes all of the steps, features, compositions and compounds referred to or indicated in the specification, individually or collectively and any and all combinations or any two or more of the steps or features.

[0032] The present invention is not to be limited in scope by the specific embodiments described herein, which are intended for the purpose of exemplification only. Functionally equivalent products, compositions and methods are clearly within the scope of the invention as described herein.

[0033] Sequence identity numbers (SEQ ID NO:) containing nucleotide and amino acid sequence information included in this specification are collected at the end of the description and have been prepared using the programme PatentIn Version 3.0. Each nucleotide or amino acid sequence is identified in the sequence listing by the numeric indicator <210> followed by the sequence identifier (e.g. <210>1, <210>2, etc.). The length, type of sequence and source organism for each nucleotide or amino acid sequence are indicated by information provided in the numeric indicator fields <211>, <212> and <213>, respectively. Nucleotide and amino acid sequences referred to in the specification are defined by the information provided in numeric indicator field <400> followed by the sequence identifier (e.g. <400>1, <400>2, etc.).

[0034] An antisense molecules nomenclature system was proposed and published to distinguish between the different antisense molecules (see Mann et al., (2002) J Gen Med 4, 644-654). This nomenclature became especially relevant when testing several slightly different antisense molecules, all directed at the same target region, as shown below:

H # A/D (x : y).

[0035] The first letter designates the species (e.g. H: human, M: murine, C: canine)

[0036] "#" designates target dystrophin exon number.

[0037] "A/D" indicates acceptor or donor splice site at the beginning and end of the exon, respectively.

[0038] (x y) represents the annealing coordinates where "-" or "+" indicate intronic or exonic sequences respectively. As an example, A(-6+18) would indicate the last 6 bases of the intron preceding the target exon and the first 18 bases of the target exon. The closest splice site would be the acceptor so these coordinates would be preceded with an "A". Describing annealing coordinates at the donor splice site could be D(+2-18) where the last 2 exonic bases and the first 18 intronic bases correspond to the annealing site of the antisense molecule. Entirely exonic annealing coordinates that would be represented by A(+65+85), that is the site between the 65th and 85th nucleotide from the start of that exon.

[0039] No admission is made that any of the references constitute prior art or are part of the common general knowledge of those working in the field to which this invention relates.

[0040] As used necessarily herein the term "derived" and "derived from" shall be taken to indicate that a specific integer may be obtained from a particular source *albeit* not directly from that source.

[0041] Throughout this specification, unless the context requires otherwise, the word "comprise", or variations such as "comprises" or "comprising", will be understood to imply the inclusion of a stated integer or group of integers but not the exclusion of any other integer or group of integers.

[0042] Other definitions for selected terms used herein may be found within the detailed description of the invention and apply throughout. Unless otherwise defined, all other scientific and technical terms used herein have the same meaning as commonly understood to one of ordinary skill in the art to which the invention belongs.

Description of the Preferred Embodiment

[0043] When antisense molecule(s) are targeted to nucleotide sequences involved in splicing in exons within pre-mRNA sequences, normal splicing of the exon may be inhibited causing the splicing machinery to by-pass the entire mutated exons from the mature mRNA. The concept of antisense oligonucleotide induced exon skipping is shown in Figure 2. In many genes, deletion of an entire exon would lead to the production of a non-functional protein through the loss of important functional domains or the disruption of the reading frame. In some proteins, however, it is possible to shorten the protein by deleting one or more exons, without disrupting the reading frame, from within the protein without seriously altering the biological activity of the protein. Typically, such proteins have a structural role and or possess functional domains at their ends. The present invention describes antisense molecules capable of binding to specified dystrophin pre-mRNA targets and re-directing processing of that gene.

Antisense Molecules

[0044] According to a first aspect of the invention, there is provided antisense molecules as defined in the claims capable of binding to a selected target to induce exon skipping. To induce exon skipping in exons of the Dystrophin gene transcript, the antisense molecules are preferably selected from the group of compounds shown in Table 1. Also described is a combination or "cocktail" of two or more antisense oligonucleotides capable of binding to a selected target to induce exon skipping.

[0045] Designing antisense molecules to completely mask consensus splice sites may not necessarily generate any skipping of the targeted exon. Furthermore, the inventors have discovered that size or length of the antisense oligonucleotide itself is not always a primary factor when designing antisense molecules. With some targets such as exon 19, antisense oligonucleotides as short as 12 bases were able to induce exon skipping, albeit not as efficiently as longer (20-31 bases) oligonucleotides. In some other targets, such as murine dystrophin exon 23, antisense oligonucleotides only 17 residues long were able to induce more efficient skipping than another overlapping compound of 25 nucleotides.

[0046] The inventors have also discovered that there does not appear to be any standard motif that can be blocked or masked by antisense molecules to redirect splicing. In some exons, such as mouse dystrophin exon 23, the donor splice site was the most amenable to target to re-direct skipping of that exon. It should be noted that designing and testing a series of exon 23 specific antisense molecules to anneal to overlapping regions of the donor splice site showed considerable variation in the efficacy of induced exon skipping. As reported in Mann et al., (2002) there was a significant variation in the efficiency of bypassing the nonsense mutation depending upon antisense oligonucleotide annealing ("Improved antisense oligonucleotide induced exon skipping in the mdx mouse model of muscular dystrophy". J Gen Med 4: 644-654). Targeting the acceptor site of exon 23 or several internal domains was not found to induce any consistent exon 23 skipping.

[0047] In other exons targeted for removal, masking the donor splice site did not induce any exon skipping. However, by directing antisense molecules to the acceptor splice site (human exon 8), strong and sustained exon skipping was induced. It should be noted that removal of human exon 8 was tightly linked with the co-removal of exon 9. There is no strong sequence homology between the exon 8 antisense oligonucleotides and corresponding regions of exon 9 so it does not appear to be a matter of cross reaction. Rather the splicing of these two exons is inextricably linked. This is not an isolated instance as the same effect is observed in canine cells where targeting exon 8 for removal also resulted in the skipping of exon 9. Targeting exon 23 for removal in the mouse dystrophin pre-mRNA also results in the frequent removal of exon 22 as well. This effect occurs in a dose dependent manner and also indicates close coordinated processing of 2 adjacent exons.

[0048] In other targeted exons, antisense molecules directed at the donor or acceptor splice sites did not induce exon skipping while annealing antisense molecules to intra-exonic regions (i.e. exon splicing enhancers within human dystrophin exon 6) was most efficient at inducing exon skipping. Some exons, both mouse and human exon 19 for example, are readily skipped by targeting antisense molecules to a variety of motifs. That is, targeted exon skipping is induced after using antisense oligonucleotides to mask donor and acceptor splice sites or exon splicing enhancers.

[0049] To identify and select antisense oligonucleotides suitable for use in the modulation of exon skipping, a nucleic acid sequence whose function is to be modulated must first be identified. This may be, for example, a gene (or mRNA transcribed form the gene) whose expression is associated with a particular disorder or disease state, or a nucleic acid molecule from an infectious agent. Within the context of the present invention, preferred target site(s) are those involved in mRNA splicing (i.e. splice donor sites, splice acceptor sites, or exonic splicing enhancer elements). Splicing branch points and exon recognition sequences or splice enhancers are also potential target sites for modulation of mRNA splicing.

[0050] Preferably, the present invention aims to provide antisense molecules capable of binding to a selected target in the dystrophin pre-mRNA to induce efficient and consistent exon skipping. Duchenne muscular dystrophy arises from mutations that preclude the synthesis of a functional dystrophin gene product. These Duchenne muscular dystrophy gene defects are typically nonsense mutations or genomic rearrangements such as deletions, duplications or microdeletions or insertions that disrupt the reading frame. As the human dystrophin gene is a large and complex gene with the 79 exons being spliced together to generate a mature mRNA with an open reading frame of approximately 11,000 bases, there are many positions where these mutations can occur. Consequently, a comprehensive antisense oligonucleotide based therapy to address many of the different disease-causing mutations in the dystrophin gene will require that many exons can be targeted for removal during the splicing process.

[0051] Within the context of the present invention, preferred target site(s) are those involved in mRNA splicing (i.e. splice donor sites, splice acceptor sites or exonic splicing enhancer elements). Splicing branch points and exon recognition sequences or splice enhancers are also potential target sites for modulation of mRNA splicing.

[0052] The oligonucleotide and the DNA or RNA are complementary to each other when a sufficient number of corresponding positions in each molecule are occupied by nucleotides which can hydrogen bond with each other. Thus, "specifically hybridisable" and "complementary" are terms which are used to indicate a sufficient degree of complementarity or precise pairing such that stable and specific binding occurs between the oligonucleotide and the DNA or RNA

target.

[0053] It is understood in the art that the sequence of an antisense molecule need not be 100% complementary to that of its target sequence to be specifically hybridisable. An antisense molecule is specifically hybridisable when binding of the compound to the target DNA or RNA molecule interferes with the normal function of the target DNA or RNA to cause a loss of utility, and there is a sufficient degree of complementarity to avoid non-specific binding of the antisense compound to non-target sequences under conditions in which specific binding is desired, i.e., under physiological conditions in the case of *in vivo* assays or therapeutic treatment, and in the case of *in vitro* assays, under conditions in which the assays are performed.

[0054] While the above method may be used to select antisense molecules capable of deleting any exon from within a protein that is capable of being shortened without affecting its biological function, the exon deletion should not lead to a reading frame shift in the shortened transcribed mRNA. Thus, if in a linear sequence of three exons the end of the first exon encodes two of three nucleotides in a codon and the next exon is deleted then the third exon in the linear sequence must start with a single nucleotide that is capable of completing the nucleotide triplet for a codon. If the third exon does not commence with a single nucleotide there will be a reading frame shift that would lead to the generation of truncated or a non-functional protein.

[0055] It will be appreciated that the codon arrangements at the end of exons in structural proteins may not always break at the end of a codon, consequently there may be a need to delete more than one exon from the pre-mRNA to ensure in-frame reading of the mRNA. In such circumstances, a plurality of antisense oligonucleotides may need to be selected by the method of the invention wherein each is directed to a different region responsible for inducing splicing in the exons that are to be deleted.

[0056] The length of an antisense molecule may vary so long as it is capable of binding selectively to the intended location within the pre-mRNA molecule. The length of such sequences can be determined in accordance with selection procedures described herein. Generally, the antisense molecule will be from about 10 nucleotides in length up to about 50 nucleotides in length. It will be appreciated however that any length of nucleotides within this range may be used in the method. Preferably, the length of the antisense molecule is between 17 to 30 nucleotides in length.

[0057] In order to determine which exons can be connected in a dystrophin gene, reference should be made to an exon boundary map. Connection of one exon with another is based on the exons possessing the same number at the 3' border as is present at the 5' border of the exon to which it is being connected. Therefore, if exon 7 were deleted, exon 6 must connect to either exons 12 or 18 to maintain the reading frame. Thus, antisense oligonucleotides would need to be selected which redirected splicing for exons 7 to 11 in the first instance or exons 7 to 17 in the second instance. Another and somewhat simpler approach to restore the reading frame around an exon 7 deletion would be to remove the two flanking exons. Induction of exons 6 and 8 skipping should result in an in-frame transcript with the splicing of exons 5 to 9. In practise however, targeting exon 8 for removal from the pre-mRNA results in the co-removal of exon 9 so the resultant transcript would have exon 5 joined to exon 10. The inclusion or exclusion of exon 9 does not alter the reading frame. Once the antisense molecules to be tested have been identified, they are prepared according to standard techniques known in the art. The most common method for producing antisense molecules is the methylation of the 2' hydroxyribose position and the incorporation of a phosphorothioate backbone produces molecules that superficially resemble RNA but that are much more resistant to nuclease degradation.

[0058] To avoid degradation of pre-mRNA during duplex formation with the antisense molecules, the antisense molecules used in the method may be adapted to minimise or prevent cleavage by endogenous RNase H. This property is highly preferred as the treatment of the RNA with the unmethylated oligonucleotides either intracellularly or in crude extracts that contain RNase H leads to degradation of the pre-mRNA: antisense oligonucleotide duplexes. Any form of modified antisense molecules that is capable of by-passing or not inducing such degradation may be used in the present method. An example, of antisense molecules which when duplexed with RNA are not cleaved by cellular RNase H is 2'-O-methyl derivatives. 2'-O-methyl-oligoribonucleotides are very stable in a cellular environment and in animal tissues, and their duplexes with RNA have higher Tm values than their ribo- or deoxyribo- counterparts.

[0059] Antisense molecules that do not activate RNase H can be made in accordance with known techniques (see, e.g., U.S. Pat. 5,149,797). Such antisense molecules, which may be deoxyribonucleotide or ribonucleotide sequences, simply contain any structural modification which sterically hinders or prevents binding of RNase H to a duplex molecule containing the oligonucleotide as one member thereof, which structural modification does not substantially hinder or disrupt duplex formation. Because the portions of the oligonucleotide involved in duplex formation are substantially different from those portions involved in RNase H binding thereto, numerous antisense molecules that do not activate RNase H are available. For example, such antisense molecules may be oligonucleotides wherein at least one, or all, of the inter-nucleotide bridging phosphate residues are modified phosphates, such as methyl phosphonates, methyl phosphorothioates, phosphoromorpholidates, phosphoropiperazidates and phosphoramidates. For example, every other one of the internucleotide bridging phosphate residues may be modified as described. In another nonlimiting example, such antisense molecules are molecules wherein at least one, or all, of the nucleotides contain a 2' lower alkyl moiety (e.g., C_1-C_4, linear or branched, saturated or unsaturated alkyl, such as methyl, ethyl, ethenyl, propyl, 1-propenyl, 2-propenyl,

and isopropyl). For example, every other one of the nucleotides may be modified as described.

[0060] While antisense oligonucleotides are a preferred form of the antisense molecules, the present invention comprehends other oligomeric antisense molecules, including but not limited to oligonucleotide mimetics such as are described below.

[0061] Specific examples of preferred antisense compounds useful in this invention include oligonucleotides containing modified backbones or non-natural inter-nucleoside linkages. As defined in this specification, oligonucleotides having modified backbones include those that retain a phosphorus atom in the backbone and those that do not have a phosphorus atom in the backbone. For the purposes of this specification, and as sometimes referenced in the art, modified oligonucleotides that do not have a phosphorus atom In their inter-nucleoside backbone can also be considered to be oligonucleosides.

[0062] In other preferred oligonucleotide mimetics, both the sugar and the inter-nucleoside linkage, i.e., the backbone; of the nucleotide units are replaced with novel groups. The base units are maintained for hybridization with an appropriate nucleic acid target compound. One such oligomeric compound, an oligonucleotide mimetic that has been shown to have excellent hybridization properties, is referred to as a peptide nucleic acid (PNA). In PNA compounds, the sugar-backbone of an oligonucleotide is replaced with an amide containing backbone, in particular an aminoethylglycine backbone. The nucleo-bases are retained and are bound directly or indirectly to aza nitrogen atoms of the amide portion of the backbone.

[0063] Modified oligonucleotides may also contain one or more substituted sugar moieties. Oligonucleotides may also include nucleobase (often referred to in the art simply as "base") modifications or substitutions. Certain nucleo-bases are particularly useful for increasing the binding affinity of the oligomeric compounds of the invention. These Include 5-substituted pyrimidines, 6-azapyrimidines and N-2, N-6 and O-6 substituted purines, including 2-aminopropyladenine, 5-propynyluracil and 5-propynylcytosine. 5-methylcytosine substitutions have been shown to increase nucleic acid duplex stability by 0.6-1.2°C and are presently preferred base substitutions, even more particularly when combined with 2'-O-methoxyethyl sugar modifications.

[0064] Another modification of the oligonucleotides of the invention involves chemically linking to the oligonucleotide one or more moieties or conjugates that enhance the activity, cellular distribution or cellular uptake of the oligonucleotide. Such moieties include but are not limited to lipid moieties such as a cholesterol moiety, cholic acid, a thioether, e.g., hexyl-S-tritylthiol, a thiocholesterol, an aliphatic chain, e.g., dodecandiol or undecyl residues, a phospholipid, e.g., di-hexadecyl-rac-glycerol or triethylammonium 1,2-di-O-hexadecyl-rac-glycero-3-H-phosphonate, a polyamine or a polyethylene glycol chain, or adamantane acetic acid, a palmityl moiety, or an octadecylamine or hexylamino-carbonyl-oxycholesterol moiety.

[0065] It is not necessary for all positions in a given compound to be uniformly modified, and in fact more than one of the aforementioned modifications may be incorporated in a single compound or even at a single nucleoside within an oligonucleotide. The present invention also includes antisense compounds that are chimeric compounds. "Chimeric" antisense compounds or "chimeras," in the context of this invention, are antisense molecules, particularly oligonucleotides, which contain two or more chemically distinct regions, each made up of at least one monomer unit, i.e., a nucleotide in the case of an oligonucleotide compound. These oligonucleotides typically contain at least one region wherein the oligonucleotide is modified so as to confer upon the Increased resistance to nuclease degradation, increased cellular uptake, and an additional region for increased binding affinity for the target nucleic acid.

Methods of Manufacturing Antisense Molecules

[0066] The antisense molecules used in accordance with this invention may be conveniently and routinely made through the well-known technique of solid phase synthesis. Equipment for such synthesis is sold by several vendors including, for example, Applied Biosystems (Foster City, Calif.). One method for synthesising oligonucleotides on a modified solid support is described in U.S. Pat. No. 4,458,066.

[0067] Any other means for such synthesis known in the art may additionally or alternatively be employed. It is well known to use similar techniques to prepare oligonucleotides such as the phosphorothioates and alkylated derivatives. In one such automated embodiment, diethyl-phosphoramidites are used as starting materials and may be synthesized as described by Beaucage, et al., (1981) Tetrahedron Letters, 22:1859-1862.

[0068] The antisense molecules of the invention are synthesised *in vitro* and do not include antisense compositions of biological origin, or genetic vector constructs designed to direct the *in vivo* synthesis of antisense molecules. The molecules of the invention may also be mixed, encapsulated, conjugated or otherwise associated with other molecules, molecule structures or mixtures of compounds, as for example, liposomes, receptor targeted molecules, oral, rectal, topical or other formulations, for assisting in uptake, distribution and/or absorption.

Therapeutic Agents

[0069] The present invention also can be used as a prophylactic or therapeutic, which may be utilised for the purpose

of treatment of a genetic disease.

[0070] Accordingly, in one embodiment the present invention provides antisense molecules that bind to a selected target in the dystrophin pre-mRNA to induce efficient and consistent exon skipping described herein in a therapeutically effective amount admixed with a pharmaceutically acceptable carrier, diluent, or excipient.

[0071] The phrase "pharmaceutically acceptable" refers to molecular entities and compositions that are physiologically tolerable and do not typically produce an allergic or similarly untoward reaction, such as gastric upset and the like, when administered to a patient. The term "carrier" refers to a diluent, adjuvant, excipient, or vehicle with which the compound is administered. Such pharmaceutical carriers can be sterile liquids, such as water and oils, including those of petroleum, animal, vegetable or synthetic origin, such as peanut oil, soybean oil, mineral oil, sesame oil and the like. Water or saline solutions and aqueous dextrose and glycerol solutions are preferably employed as carriers, particularly for injectable solution. Suitable pharmaceutical carriers are described in Martin, Remington's Pharmaceutical Sciences, 18th Ed., Mack Publishing Co., Easton, PA, (1990).

[0072] In a more specific form of the invention there are provided pharmaceutical compositions comprising therapeutically effective amounts of an antisense molecule together with pharmaceutically acceptable diluents, preservatives, solubilizers, emulsifiers, adjuvants and/or carriers. Such compositions include diluents of various buffer content (*e.g.*, Tris-HCl, acetate, phosphate), pH and ionic strength and additives such as detergents and solubilizing agents (e.g., Tween 80, Polysorbate 80), anti-oxidants (*e.g.*, ascorbic acid, sodium metabisulfite), preservatives (*e.g.,* Thimersol, benzyl alcohol) and bulking substances (*e.g.,* lactose, mannitol). The material may be incorporated into particulate preparations of polymeric compounds such as polylactic acid, polyglycolic acid, etc. or into liposomes. Hylauronic acid may also be used. Such compositions may influence the physical state, stability, rate of *in vivo* release, and rate of *in vivo* clearance of the present proteins and derivatives. *See, e.g.,* Martin, Remington's Pharmaceutical Sciences, 18th Ed. (1990, Mack Publishing Co., Easton, PA 18042) pages 1435-1712. The compositions may be prepared in liquid form, or may be in dried powder, such as lyophilised form.

[0073] It will be appreciated that pharmaceutical compositions provided according to the present invention may be administered by any means known in the art. Preferably, the pharmaceutical compositions for administration are administered by injection, orally, or by the pulmonary, or nasal route. The antisense molecules are more preferably delivered by intravenous, intra-arterial, intraperitoneal, intramuscular, or subcutaneous routes of administration.

Antisense molecule based therapy

[0074] Also addressed by the present invention is the use of antisense molecules of the present invention, for manufacture of a medicament for modulation of a genetic disease.

[0075] The delivery of a therapeutically useful amount of antisense molecules may be achieved by methods previously published. For example, intracellular delivery of the antisense molecule may be via a composition comprising an admixture of the antisense molecule and an effective amount of a block copolymer. An example of this method is described in US patent application US 20040248833.

[0076] Other methods of delivery of antisense molecules to the nucleus are described in Mann CJ et al., (2001) ["Antisense-induced exon skipping and the synthesis of dystrophin in the mdx mouse". Proc., Natl. Acad. Science, 98(1) 42-47] and in Gebski et al., (2003). Human Molecular Genetics, 12(15): 1801-1811.

[0077] A method for introducing a nucleic acid molecule into a cell by way of an expression vector either as naked DNA or complexed to lipid carriers, is described in US patent US 6,806,084.

[0078] It may be desirable to deliver the antisense molecule in a colloidal dispersion system. Colloidal dispersion systems include macromolecule complexes, nanocapsules, microspheres, beads, and lipid-based systems including oil-in-water emulsions, micelles, mixed micelles, and liposomes or liposome formulations.

[0079] Liposomes are artificial membrane vesicles which are useful as delivery vehicles *in vitro* and *in vivo*. These formulations may have net cationic, anionic or neutral charge characteristics and are useful characteristics with *in vitro, in vivo* and *ex vivo* delivery methods. It has been shown that large unilamellar vesicles (LUV), which range in size from 0.2-4.0 .PHI.m can encapsulate a substantial percentage of an aqueous buffer containing large macromolecules. RNA, and DNA can be encapsulated within the aqueous interior and be delivered to cells in a biologically active form (Fraley, et al., Trends Biochem. Sci., 6:77, 1981).

[0080] In order for a liposome to be an efficient gene transfer vehicle, the following characteristics should be present: (1) encapsulation of the antisense molecule of interest at high efficiency while not compromising their biological activity; (2) preferential and substantial binding to a target cell in comparison to non-target cells; (3) delivery of the aqueous contents of the vesicle to the target cell cytoplasm at high efficiency; and (4) accurate and effective expression of genetic information (Mannino, et al., Biotechniques, 6:682, 1988).

[0081] The composition of the liposome is usually a combination of phospholipids, particularly high-phase-transition-temperature phospholipids, usually in combination with steroids, especially cholesterol. Other phospholipids or other lipids may also be used. The physical characteristics of liposomes depend on pH, ionic strength, and the presence of

divalent cations.

[0082] Alternatively, the antisense construct may be combined with other pharmaceutically acceptable carriers or diluents to produce a pharmaceutical composition. Suitable carriers and diluents include isotonic saline solutions, for example phosphate-buffered saline. The composition may be formulated for parenteral, intramuscular, intravenous, subcutaneous, intraocular, oral or transdermal administration.

[0083] The routes of administration described are intended only as a guide since a skilled practitioner will be able to determine readily the optimum route of administration and any dosage for any particular animal and condition. Multiple approaches for introducing functional new genetic material into cells, both in vitro and in vivo have been attempted (Friedmann (1989) Science, 244:1275-1280).

[0084] These approaches include integration of the gene to be expressed into modified retroviruses (Friedmann (1989) supra; Rosenberg (1991) Cancer Research 51(18), suppl.: 5074S-5079S); integration into non-retrovirus vectors (Rosenfeld, et al. (1992) Cell, 68:143-155; Rosenfeld, et al. (1991) Science, 252:431-434); or delivery of a transgene linked to a heterologous promoter-enhancer element via liposomes (Friedmann (1989), supra; Brigham, et al. (1989) Am. J. Med. Sci.. 298:278-281; Nabel, et al. (1990) Science, 249:1285-1288; Hazinski, et al. (1991) Am. J. Resp. Cell Molec. Biol., 4:206-209; and Wang and Huang (1987) Proc. Natl. Acad. Sci. (USA), 84:7851-7855); coupled to ligand-specific, cation-based transport systems (Wu and Wu (1988) J. Biol. Chem., 263:14621-14624) or the use of naked DNA, expression vectors (Nabel et al. (1990), supra); wolff et al. (1990) Science, 247:1465-1468). Direct injection of transgenes into tissue produces only localized expression (Rosenfeld (1992) supra); Rosenfeld et al. (1991) supra; Brigham et al. (1989) supra; Nabel (1990) supra; and Hazinski et al. (1991) supra). The Brigham et al. group (Am. J. Med. Sci. (1989) 298:278-281 and Clinical Research (1991) 39 (abstract)) have reported in vivo transfection only of lungs of mice following either intravenous or intratracheal administration of a DNA liposome complex. An example of a review article of human gene therapy procedures is: Anderson, Science (1992) 256:808-813.

[0085] The antisense molecules of the invention encompass any pharmaceutically acceptable salts, esters, or salts of such esters, or any other compound which, upon administration to an animal including a human, is capable of providing (directly or indirectly) the biologically active metabolite or residue thereof. Accordingly, for example, the disclosure is also drawn to prodrugs and pharmaceutically acceptable salts of the compounds of the invention, pharmaceutically acceptable salts of such pro-drugs, and other bioequivalents.

[0086] The term "pharmaceutically acceptable salts" refers to physiologically and pharmaceutically acceptable salts of the compounds of the invention: i.e., salts that retain the desired biological activity of the parent compound and do not impart undesired toxicological effects thereto.

[0087] For oligonucleotides, preferred examples of pharmaceutically acceptable salts include but are not limited to (a) salts formed with cations such as sodium, potassium, ammonium, magnesium, calcium, polyamines such as spermine and spermidine, etc.; (b) acid addition salts formed with inorganic acids, for example hydrochloric acid, hydrobromic acid, sulfuric acid, phosphoric acid, nitric acid and the like; (c) salts formed with organic acids such as, for example, acetic acid, oxalic acid, tartaric acid, succinic acid, maleic acid, fumaric acid, gluconic acid, citric acid, malic acid, ascorbic acid, benzoic acid, tannic acid, palmitic acid, alginic acid, polyglutamic acid, naphthalenesulfonic acid, methanesulfonic acid, p-toluenesulfonic acid, naphthalenedisulfonic acid, polygalacturonic acid, and the like; and (d) salts formed from elemental anions such as chlorine, bromine, and iodine. The pharmaceutical compositions of the present invention may be administered in a number of ways depending upon whether local or systemic treatment is desired and upon the area to be treated. Administration may be topical (including ophthalmic and to mucous membranes including rectal delivery), pulmonary, e.g., by inhalation or insufflation of powders or aerosols, (including by nebulizer, intratracheal, intranasal, epidermal and transdermal), oral or parenteral. Parenteral administration includes intravenous, intra-arterial, subcutaneous, intraperitoneal or intramuscular injection or infusion; or intracranial, e.g., intrathecal or intraventricular, administration. Oligonucleotides with at least one 2'-O-methoxyethyl modification are believed to be particularly useful for oral administration.

[0088] The pharmaceutical formulations of the present invention, which may conveniently be presented in unit dosage form, may be prepared according to conventional techniques well known in the pharmaceutical industry. Such techniques include the step of bringing into association the active ingredients with the pharmaceutical carrier(s) or excipient(s). In general the formulations are prepared by uniformly and intimately bringing into association the active ingredients with liquid carriers or finely divided solid carriers or both, and then, if necessary, shaping the product.

Kits of the Invention

[0089] The invention also provides kits for treatment of a patient with a genetic disease which kit comprises at least an antisense molecule, packaged in a suitable container, together with instructions for its use.

[0090] In a preferred embodiment, the kits will contain at least one antisense molecule as defined in the claims. The kits may also contain peripheral reagents such as buffers, stabilizers, etc.

[0091] Those of ordinary skill in the field should appreciate that applications of the above method has wide application

for identifying antisense molecules suitable for use in the treatment of many other diseases.

EXAMPLES

[0092] The following Examples serve to more fully describe the manner of using the above-described invention, as well as to set forth the best modes contemplated for carrying out various aspects of the invention. It is understood that these Examples in no way serve to limit the true scope of this invention, but rather are presented for illustrative purposes.

[0093] Methods of molecular cloning, immunology and protein chemistry, which are not explicitly described in the following examples, are reported in the literature and are known by those skilled in the art. General texts that described conventional molecular biology, microbiology, and recombinant DNA techniques within the skill of the art, included, for example: Sambrook et al., Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1989); Glover ed., DNA Cloning: A Practical Approach, Volumes I and II, MRL Press, Ltd., Oxford, U.K. (1985); and Ausubel, F., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., Struhl, K. Current Protocols in Molecular Biology. Greene Publishing Associates/Wiley Intersciences, New York (2002).

Determining Induced Exon Skipping in Human Muscle Cells

[0094] Attempts by the inventors to develop a rational approach in antisense molecules design were not completely successful as there did not appear to be a consistent trend that could be applied to all exons. As such, the identification of the most effective and therefore most therapeutic antisense molecules compounds has been the result of empirical studies.

[0095] These empirical studies involved the use of computer programs to identify motifs potentially involved in the splicing process. Other computer programs were also used to identify regions of the pre-mRNA which may not have had extensive secondary structure and therefore potential sites for annealing of antisense molecules. Neither of these approaches proved completely reliable in designing antisense oligonucleotides for reliable and efficient induction of exon skipping.

[0096] Annealing sites on the human dystrophin pre-mRNA were selected for examination, initially based upon known or predicted motifs or regions involved in splicing. 2OMe antisense oligonucleotides were designed to be complementary to the target sequences under investigation and were synthesised on an Expedite 8909 Nucleic Acid Synthesiser. Upon completion of synthesis, the oligonucleotides were cleaved from the support column and de-protected in ammonium hydroxide before being desalted. The quality of the oligonucleotide synthesis was monitored by the intensity of the trityl signals upon each deprotection step during the synthesis as detected in the synthesis log. The concentration of the antisense oligonucleotide was estimated by measuring the absorbance of a diluted aliquot at 260nm.

[0097] Specified amounts of the antisense molecules were then tested for their ability to induce exon skipping in an *in vitro* assay, as described below.

[0098] Briefly, normal primary myoblast cultures were prepared from human muscle biopsies obtained after informed consent. The cells were propagated and allowed to differentiate into myotubes using standard culturing techniques. The cells were then transfected with the antisense oligonucleotides by delivery of the oligonucleotides to the cells as cationic lipoplexes, mixtures of antisense molecules or cationic liposome preparations.

[0099] The cells were then allowed to grow for another 24 hours, after which total RNA was extracted and molecular analysis commenced. Reverse transcriptase amplification (RT-PCR) was undertaken to study the targeted regions of the dystrophin pre-mRNA or induced exonic re-arrangements.

[0100] For example, in the testing of an antisense molecule for inducing exon 19 skipping the RT-PCR test scanned several exons to detect involvement of any adjacent exons. For example, when inducing skipping of exon 19, RT-PCR was carried out with primers that amplified across exons 17 and 21. Amplifications of even larger products in this area (i.e. exons 13-26) were also carried out to ensure that there was minimal amplification bias for the shorter induced skipped transcript. Shorter or exon skipped products tend to be amplified more efficiently and may bias the estimated of the normal and induced transcript.

[0101] The sizes of the amplification reaction products were estimated on an agarose gel and compared against appropriate size standards. The final confirmation of identity of these products was carried out by direct DNA sequencing to establish that the correct or expected exon junctions have been maintained.

[0102] Once efficient exon skipping had been induced with one antisense molecule, subsequent overlapping antisense molecules may be synthesized and then evaluated in the assay as described above. Our definition of an efficient antisense molecule is one that induces strong and sustained exon skipping at transfection concentrations in the order of 300 nM or less.

Antisense Oligonucleotides Directed at Exon 53

[0103] Antisense oligonucleotides directed at exon 53 were prepared and tested for their ability to induce exon skipping in human muscle cells using similar methods as described above.

[0104] Figure 3 also shows antisense molecule H53A(+39+69) [SEQ ID NO:193] directed at exon 53 acceptor splice site. This antisense oligonucleotide was able to induce exon 53 skipping at 5, 100, 300 and 600 nM. A "cocktail" of three exon 53 antisense oligonucleotides:- H53D(+23+47) [SEQ ID NO:195], H53A(+150+175) [SEQ ID NO:196] and H53A(+14-07) [SEQ ID NO:194], were also tested, as shown in Figure 3 and exhibited an ability to induce exon skipping.

[0105] Table 2 below includes other antisense molecules tested at a concentration range of 50, 100, 300 and 600 nM. These antisense molecules showed varying ability to induce exon 53 skipping. Antisense molecule H53A(+39+69) [SEQ ID NO:193] induced the strongest exon 53 skipping.

Table 2

Antisense oligonucleotide name	Sequence	Ability to induce skipping
H53A(+45+69)	CAU UCA ACU GUU GCC UCC GGU UCU G	Faint skipping at 50 nM
H53A(+39+62)	CUG UUG CCU CCG GUU CUG AAG GUG	Faint skipping at 50 nM
H53A(+39+69)	CAU UCA ACU GUU GCC UCC GGU UCU GAA GGU G	Strong skipping to 50 nM
H53D(+14-07)	UAC UAA CCU UGG UUU CUG UGA	Very faint skipping to 50 nM
H53A(+23+47)	CUG AAG GUG UUC UUG UAC UUC AUC C	Very faint skipping to 50 nM
H53A(+150+176)	UGU AUA GGG ACC CUC CUU CCA UGA CUC	Very faint skipping to 50 nM
H53D(+20-05)	CUA ACC UUG GUU UCU GUG AUU UUC U	Not made yet
H53D(+09-18)	GGU AUC UUU GAU ACU AAC CUU GGU UUC	Faint at 600 nM
H53A(-12+10)	AUU CUU UCA ACU AGA AUA AAA G	No skipping
H53A(-07+18)	GAU UCU GAA UUC UUU CAA CUA GAA U	No skipping
H53A(+07+26)	AUC CCA CUG AUU CUG AAU UC	No skipping
H53A(+124+145)	UUG GCU CUG GCC UGU CCU AAG A	No skipping

[0106] The invention is defined with reference to the following clauses:

Clause 1: An antisense molecule capable of binding to a selected target site to induce exon skipping in the dystrophin gene, as set forth in SEQ ID NO: 1 to 202.

Clause 2: An antisense molecule according to clause 1 capable of inducing exon skipping in exons 3, 4, 8, 10 to 16, 19 to 40, 42 to 44, 46, 47 and 50 to 53 of the dystrophin gene.

Clause 3: A combination of two or more antisense molecules according to clause 1 or clause 2 capable of binding to a selected target to induce exon skipping in the dystrophin gene.

Clause 4: A combination or two or more antisense molecules according to clause 3 selected from Table 1B.

Clause 5: A combination of two or more antisense molecules according to clause 1 or clause 2 joined together to form a "weasel", wherein said weasel is capable of binding to a selected target to induce exon skipping in the dystrophin gene.

Clause 6: A combination of two or more antisense molecules according to clause 5 selected from Table 1C.

Clause 7: The antisense molecule according to anyone of clauses 1 to 6, capable of binding to a selected target site, wherein the target site is an mRNA splicing site selected from a splicer donor site, splice acceptor sites or exonic splicing enhancer elements.

Clause 8: A method of treating muscular dystrophy in a patient comprising administering to the patient a composition comprising an antisense molecule according to anyone of clauses 1 to 6.

Clause 9: A pharmaceutical or therapeutic composition for the treatment of muscular dystrophy in a patient comprising (a) at least an antisense molecule according to anyone of clause 1 to 6, and (b) one or more pharmaceutically acceptable carriers and/or diluents.

Clause 10: The composition according to clause 9, comprising about 20 nM to 600 nM of the antisense molecule.

Clause 11. The use of an antisense molecule according to anyone of clauses 1 to 6 for the manufacture of a medicament for modulation of muscular dystrophy.

Clause 12. An antisense molecule according to anyone of clauses 1 to 6 for use in antisense molecule based therapy.

Clause 13. An antisense molecule according to anyone of clauses 1 to 6 as herein before described with reference to the examples.

Clause 14. A kit comprising at least one antisense molecule according to anyone of clauses 1 to 6, a suitable carrier and instructions for its use.

SEQUENCE LISTING

[0107]

<110> The University of Western Australia

<120> Antisense Oligonucleotides for Inducing Exon Skipping and Methods of Use Thereof

<130> AHB/FP6689632

<140> EP10004274.6
<141> 2005-06-28

<150> EP05754344.9
<151> 2005-06-28

<150> PCT/AU2005/000943
<151> 2005-06-28

<150> AU20040903474
<151> 2004-06-28

<160> 212

<170> PatentIn version 3.3

<210> 1
<211> 24
<212> RNA
<213> Human

<400> 1

gauagguggu aucaacaucu guaa 24

<210> 2
<211> 21
<212> RNA
<213> Human

<400> 2
gauagguggu aucaacaucu g 21

<210> 3
<211> 25
<212> RNA
<213> Human

<400> 3
gauagguggu aucaacaucu guaag 25

<210> 4
<211> 20
<212> RNA
<213> Human

<400> 4
ggugguauca acaucuguaa 20

<210> 5
<211> 20
<212> RNA
<213> Human

<400> 5
guaucaacau cuguaagcac 20

<210> 6
<211> 23
<212> RNA
<213> Human

<400> 6
ugcauguucc agucguugug ugg 23

<210> 7
<211> 25
<212> RNA
<213> Human

<400> 7
cacuauucca gucaaauagg ucugg 25

<210> 8
<211> 25
<212> RNA
<213> Human

<400> 8
auuuaccaac cuucaggauc gagua 25

<210> 9
<211> 21
<212> RNA
<213> Human

<400> 9
ggccuaaaac acauacacau a 21

<210> 10
<211> 20
<212> RNA
<213> Canine

<400> 10
cauuuuugac cuacaugugg 20

<210> 11
<211> 20
<212> RNA
<213> Canine

<400> 11
uuugaccuac auguggaaag 20

<210> 12
<211> 26
<212> RNA
<213> Canine

<400> 12
uacauuuuug accuacaugu ggaaag 26

<210> 13
<211> 22
<212> RNA
<213> Canine

<400> 13
auuuuugacc uacaugggaa ag 22

<210> 14
<211> 23
<212> RNA
<213> Canine

<400> 14
uacgaguuga uugucggacc cag 23

<210> 15
<211> 25
<212> RNA
<213> Canine

<400> 15
guggucuccu uaccuaugac ugugg 25

<210> 16
<211> 17

<212> RNA
<213> Canine

<400> 16
ggucuccuua ccuauga 17

<210> 17
<211> 24
<212> RNA
<213> Human

<400> 17
ugucucagua aucuucuuac cuau 24

<210> 18
<211> 24
<212> RNA
<213> Human

<400> 18
ucuuaccuau gacuauggau gaga 24

<210> 19
<211> 20
<212> RNA
<213> Human

<400> 19
gcaugaacuc uuguggaucc 20

<210> 20
<211> 20
<212> RNA
<213> Human

<400> 20
ccaggguacu acuuacauua 20

<210> 21
<211> 21
<212> RNA
<213> Human

<400> 21
aucgugugc acagcaucca g 21

<210> 22
<211> 30
<212> RNA
<213> Human

<400> 22
uguucagggc augaacucuu guggauccuu 30

<210> 23
<211> 31
<212> RNA
<213> Human

<400> 23
uaggaggcgc cucccauccu guaggucacu g 31

<210> 24
<211> 31
<212> RNA
<213> Human

<400> 24
aggucuagga ggcgccuccc auccuguagg u 31

<210> 25
<211> 25
<212> RNA
<213> Human

<400> 25
gcgccuccca uccuguaggu cacug 25

<210> 26
<211> 26
<212> RNA
<213> Human

<400> 26
cuucgaggag gucuaggagg cgccuc 26

<210> 27
<211> 21
<212> RNA
<213> Human

<400> 27
cucccauccu guaggucacu g 21

<210> 28
<211> 22
<212> RNA
<213> Human

<400> 28
uaccaguuuu ugcccuguca gg 22

<210> 29
<211> 26
<212> RNA
<213> Human

<400> 29
ucaauaugcu gcuucccaaa cugaaa 26

<210> 30
<211> 25
<212> RNA
<213> Human

<400> 30
cuaggaggcg ccucccaucc uguag 25

<210> 31
<211> 31
<212> RNA
<213> Human

<400> 31
uuaugauuuc caucuacgau gucaguacuu c 31

<210> 32
<211> 31
<212> RNA
<213> Human

<400> 32
cuuaccugcc aguggaggau uauauuccaa a 31

<210> 33
<211> 25
<212> RNA
<213> Human

<400> 33
caucaggauu cuuaccugcc agugg 25

<210> 34
<211> 25
<212> RNA
<213> Human

<400> 34
cgaugucagu acuuccaaua uucac 25

<210> 35
<211> 18
<212> RNA
<213> Human

<400> 35
accauucauc aggauucu 18

<210> 36
<211> 18
<212> RNA
<213> Human

<400> 36
accugccagu ggaggauu 18

<210> 37
<211> 27
<212> RNA
<213> Human

<400> 37
ccaauauuca cuaaaucaac cuguuaa 27

<210> 38
<211> 30

<212> RNA
<213> Human

<400> 38
caggauucuu accugccagu ggaggauuau 30

<210> 39
<211> 31
<212> RNA
<213> Human

<400> 39
acgaugucag uacuuccaau auucacuaaa u 31

<210> 40
<211> 31
<212> RNA
<213> Human

<400> 40
auuuccaucu acgaugucag uacuuccaau a 31

<210> 41
<211> 21
<212> RNA
<213> Human

<400> 41
caggagcuuc caaaugcugc a 21

<210> 42
<211> 29
<212> RNA
<213> Human

<400> 42
cuugucuuca ggagcuucca aaugcugca 29

<210> 43
<211> 22
<212> RNA
<213> Human

<400> 43
uccucagcag aaagaagcca cg 22

<210> 44
<211> 20
<212> RNA
<213> Human

<400> 44
uuagaaaucu cuccuugugc 20

<210> 45
<211> 20
<212> RNA
<213> Human

<400> 45
uaaauugggu guuacacaau 20

<210> 46
<211> 24
<212> RNA
<213> Human

<400> 46
cccugaggca uucccaucuu gaau 24

<210> 47
<211> 20
<212> RNA
<213> Human

<400> 47
aggacuuacu ugcuuuguuu 20

<210> 48
<211> 23
<212> RNA
<213> Human

<400> 48
cuugaauuua ggagauucau cug 23

<210> 49
<211> 23
<212> RNA
<213> Human

<400> 49
caucuucuga uaauuuuccu guu 23

<210> 50
<211> 24
<212> RNA
<213> Human

<400> 50
ucuucuguuu uuguuagcca guca 24

<210> 51
<211> 20
<212> RNA
<213> Human

<400> 51
ucuauguaaa cugaaaauuu 20

<210> 52
<211> 20
<212> RNA
<213> Human

<400> 52
uucuggagau ccauuaaaac 20

<210> 53
<211> 24
<212> RNA
<213> Human

<400> 53
cagcaguugc gugaucucca cuag 24

<210> 54
<211> 21
<212> RNA
<213> Human

<400> 54
uucaucaacu accaccacca u 21

<210> 55
<211> 25
<212> RNA
<213> Human

<400> 55
cuaagcaaaa uaaucugacc uuaag 25

<210> 56
<211> 28
<212> RNA
<213> Human

<400> 56
cuuguaaaag aacccagcgg ucuucugu 28

<210> 57
<211> 22
<212> RNA
<213> Human

<400> 57
caucuacaga uguuugccca uc 22

<210> 58
<211> 23
<212> RNA
<213> Human

<400> 58
gaaggauguc uuguaaaaga acc 23

<210> 59
<211> 20
<212> RNA
<213> Human

<400> 59
accuguucuu caguaagacg 20

<210> 60
<211> 24

<212> RNA
<213> Human

<400> 60
caugacacac cuguucuuca guaa 24

<210> 61
<211> 20
<212> RNA
<213> Human

<400> 61
cauuugagaa ggaugucuug 20

<210> 62
<211> 24
<212> RNA
<213> Human

<400> 62
aucucccaau accuggagaa gaga 24

<210> 63
<211> 31
<212> RNA
<213> Human

<400> 63
gccaugcacu aaaaaggcac ugcaagacau u 31

<210> 64
<211> 24
<212> RNA
<213> Human

<400> 64
ucuuuaaagc caguugugug aauc 24

<210> 65
<211> 21
<212> RNA
<213> Human

<400> 65
uuucugaaag ccaugcacua a 21

<210> 66
<211> 25
<212> RNA
<213> Human

<400> 66
guacauacgg ccaguuuuug aagac 25

<210> 67
<211> 31
<212> RNA
<213> Human

<400> 67
cuagauccgc uuuuaaaacc uguuaaaaca a 31

<210> 68
<211> 31
<212> RNA
<213> Human

<400> 68
ucuuuucuag auccgcuuuu aaaaccuguu a 31

<210> 69
<211> 25
<212> RNA
<213> Human

<400> 69
cuagauccgc uuuuaaaacc uguua 25

<210> 70
<211> 23
<212> RNA
<213> Human

<400> 70
ccgucuucug ggucacugac uua 23

<210> 71
<211> 26
<212> RNA
<213> Human

<400> 71
cuagauccgc uuuuaaaacc uguuaa 26

<210> 72
<211> 20
<212> MRNA
<213> Human

<400> 72
ccgcuuuuaa aaccuguuaa 20

<210> 73
<211> 26
<212> RNA
<213> Human

<400> 73
uggauugcuu uuucuuuucu agaucc 26

<210> 74
<211> 25
<212> RNA
<213> Human

<400> 74
caugcuuccg ucuucugggu cacug 25

<210> 75
<211> 23
<212> RNA
<213> Human

<400> 75
gaucuuguuu gagugaauac agu 23

<210> 76
<211> 22
<212> RNA
<213> Human

<400> 76
guuauccagc caugcuuccg uc 22

<210> 77
<211> 25
<212> RNA
<213> Human

<400> 77
ugauaauugg uaucacuaac cugug 25

<210> 78
<211> 22
<212> RNA
<213> Human

<400> 78
guaucacuaa ccugugcugu ac 22

<210> 79
<211> 20
<212> RNA
<213> Human

<400> 79
cagcaguagu ugucaucugc 20

<210> 80
<211> 31
<212> RNA
<213> Human

<400> 80
gccugagcug aucugcuggc aucuugcagu u 31

<210> 81
<211> 28
<212> RNA
<213> Human

<400> 81
cuggcagaau ucgauccacc ggcuguuc 28

<210> 82
<211> 22

<212> RNA
<213> Human

<400> 82
cagcaguagu ugucaucugc uc 22

<210> 83
<211> 19
<212> RNA
<213> Human

<400> 83
ugaugggggug gugggguugg 19

<210> 84
<211> 25
<212> RNA
<213> Human

<400> 84
aucugcauua acacccucua gaaag 25

<210> 85
<211> 24
<212> RNA
<213> Human

<400> 85
ccggcuguuc aguuguucug aggc 24

<210> 86
<211> 28
<212> RNA
<213> Human

<400> 86
aucugcauua acacccucua gaaagaaa 28

<210> 87
<211> 28
<212> RNA
<213> Human

<400> 87
gaaggagaag agauucuuac cuuacaaa 28

<210> 88
<211> 20
<212> RNA
<213> Human

<400> 88
auucgaucca ccggcuguuc 20

<210> 89
<211> 19
<212> RNA
<213> Human

<400> 89
cugcuggcau cuugcaguu 19

<210> 90
<211> 22
<212> DNA
<213> Human

<400> 90
gccgguugac uucauccugu gc 22

<210> 91
<211> 22
<212> RNA
<213> Human

<400> 91
cugcauccag gaacaugggu cc 22

<210> 92
<211> 23
<212> RNA
<213> Human

<400> 92
gucugcaucc aggaacaugg guc 23

<210> 93
<211> 24
<212> DNA
<213> Human

<400> 93
guugaagauc ugauagccgg uuga 24

<210> 94
<211> 24
<212> RNA
<213> Human

<400> 94
uacuuacugu cuguagcucu uucu 24

<210> 95
<211> 24
<212> RNA
<213> Human

<400> 95
cacucauggu cuccugauag cgca 24

<210> 96
<211> 22
<212> RNA
<213> Human

<400> 96
cugcaauucc ccgagucucu gc 22

<210> 97
<211> 23
<212> RNA
<213> Human

<400> 97
acugcuggac ccauguccug aug 23

<210> 98
<211> 21
<212> RNA
<213> Human

<400> 98
cuaaguugag guauggagag u 21

<210> 99
<211> 23
<212> RNA
<213> Human

<400> 99
uauucacaga ccugcaauuc ccc 23

<210> 100
<211> 26
<212> RNA
<213> Human

<400> 100
acaguggugc ugagauagua uaggcc 26

<210> 101
<211> 22
<212> RNA
<213> Human

<400> 101
uaggccacuu uguugcucuu gc 22

<210> 102
<211> 19
<212> RNA
<213> Human

<400> 102
uucagagggc gcuuucuuc 19

<210> 103
<211> 23
<212> DNA
<213> Human

<400> 103
gggcaggcca uuccuccuuc aga 23

<210> 104
<211> 24

<212> RNA
<213> Human

<400> 104
ucuucagggu uuguauguga uucu 24

<210> 105
<211> 27
<212> RNA
<213> Human

<400> 105
cugggcugaa uugucugaau aucacug 27

<210> 106
<211> 26
<212> RNA
<213> Human

<400> 106
cuguuggcac augugauccc acugag 26

<210> 107
<211> 24
<212> RNA
<213> Human

<400> 107
gucuauaccu guuggcacau guga 24

<210> 108
<211> 25
<212> RNA
<213> Human

<400> 108
ugcuuucugu aauucaucug gaguu 25

<210> 109
<211> 26
<212> RNA
<213> Human

<400> 109
ccuccuuucu ggcauagacc uuccac 26

<210> 110
<211> 25
<212> RNA
<213> Human

<400> 110
ugugucaucc auucgugcau cucug 25

<210> 111
<211> 25
<212> RNA
<213> Human

<400> 111
uuaaggccuc uugugcuaca ggugg 25

<210> 112
<211> 23
<212> MRNA
<213> Human

<400> 112
gggccucuuc uuuagcucuc uga 23

<210> 113
<211> 22
<212> RNA
<213> Human

<400> 113
gacuuccaaa gucuugcauu uc 22

<210> 114
<211> 24
<212> RNA
<213> Human

<400> 114
gccaacaugc ccaaacuucc uaag 24

<210> 115
<211> 26
<212> RNA
<213> Human

<400> 115
cagagauuuc cucagcuccg ccagga 26

<210> 116
<211> 21
<212> RNA
<213> Human

<400> 116
cuuacaucua gcaccucaga g 21

<210> 117
<211> 25
<212> RNA
<213> Human

<400> 117
uccgccaucu guuagggucu gugcc 25

<210> 118
<211> 25
<212> RNA
<213> Human

<400> 118
auuuggguua uccucugaau gucgc 25

<210> 119
<211> 22
<212> RNA
<213> Human

<400> 119
cauaccucuu cauguaguuc uc 22

<210> 120
<211> 31
<212> RNA
<213> Human

<400> 120
cauuugagcu gcguccaccu ugucgucugu g 31

<210> 121
<211> 26
<212> RNA
<213> Human

<400> 121
uccugggcag acuggaugcu cuguuc 26

<210> 122
<211> 23
<212> RNA
<213> Human

<400> 122
uugccugggc uuccugaggc auu 23

<210> 123
<211> 24
<212> RNA
<213> Human

<400> 123
uucugaaaua acauauaccu gugc 24

<210> 124
<211> 25
<212> RNA
<213> Human

<400> 124
uaguuucuga aauaacauau accug 25

<210> 125
<211> 21
<212> RNA
<213> Human

<400> 125
gacuugucaa aucagauugg a 21

<210> 126
<211> 24

<212> RNA
<213> Human

<400> 126
guuucugaaa uaacauauac cugu 24

<210> 127
<211> 20
<212> RNA
<213> Human

<400> 127
caccagaaau acauaccaca 20

<210> 128
<211> 20
<212> RNA
<213> Human

<400> 128
caaugauuua gcugugacug 20

<210> 129
<211> 23
<212> RNA
<213> Human

<400> 129
cgaaacuuca uggagacauc uug 23

<210> 130
<211> 25
<212> RNA
<213> Human

<400> 130
cuuguagacg cugcucaaaa uuggc 25

<210> 131
<211> 20
<212> RNA
<213> Human

<400> 131
caugcacaca ccuuugcucc 20

<210> 132
<211> 24
<212> RNA
<213> Human

<400> 132
ucuguacaau cugacgucca gucu 24

<210> 133
<211> 27
<212> RNA
<213> Human

<400> 133
gucuuuauca ccauuuccac uucagac 27

<210> 134
<211> 25
<212> RNA
<213> Human

<400> 134
ccgucugcuu uuucuguaca aucug 25

<210> 135
<211> 22
<212> RNA
<213> Human

<400> 135
uccauaucug uagcugccag cc 22

<210> 136
<211> 23
<212> RNA
<213> Human

<400> 136
ccaggcaacu ucagaaucca aau 23

<210> 137
<211> 30
<212> RNA
<213> Human

<400> 137
uuucuguuac cugaaaagaa uuauaaugaa 30

<210> 138
<211> 25
<212> RNA
<213> Human

<400> 138
cauucauuuc cuuucgcauc uuacg 25

<210> 139
<211> 26
<212> RNA
<213> Human

<400> 139
ugaucucuuu gucaauucca uaucug 26

<210> 140
<211> 27
<212> RNA
<213> Human

<400> 140
uucagugaua uagguuuuac cuuuccc 27

<210> 141
<211> 26
<212> DNA
<213> Human

<400> 141
cuguagcugc cagccauucu gucaag 26

<210> 142
<211> 21
<212> RNA
<213> Human

<400> 142
ucuucugcuc gggaggugac a 21

<210> 143
<211> 20
<212> RNA
<213> Human

<400> 143
ccaguuacua uucagaagac 20

<210> 144
<211> 20
<212> RNA
<213> Human

<400> 144
ucuucaggug caccuucugu 20

<210> 145
<211> 25
<212> RNA
<213> Human

<400> 145
ugugaugugg uccacauucu gguca 25

<210> 146
<211> 20
<212> RNA
<213> Human

<400> 146
ccauguguuu cugguauucc 20

<210> 147
<211> 25
<212> RNA
<213> Human

<400> 147
cguguagagu ccaccuuugg gcgua 25

<210> 148
<211> 24

<212> RNA
<213> Human

<400> 148
uacuaauuuc cugcaguggu cacc 24

<210> 149
<211> 24
<212> RNA
<213> Human

<400> 149
uucuguguga aauggcugca aauc 24

<210> 150
<211> 20
<212> RNA
<213> Human

<400> 150
ccuucaaagg aauggaggcc 20

<210> 151
<211> 25
<212> RNA
<213> Human

<400> 151
ugcugaauuu cagccuccag ugguu 25

<210> 152
<211> 25
<212> RNA
<213> Human

<400> 152
ugaagucuuc cucuuucaga uucac 25

<210> 153
<211> 24
<212> RNA
<213> Human

<400> 153
cuggcuuucu cucaucugug auuc 24

<210> 154
<211> 20
<212> RNA
<213> Human

<400> 154
guuguaaguu gucuccucuu 20

<210> 155
<211> 20
<212> RNA
<213> Human

<400> 155
uugucuguaa cagcugcugu 20

<210> 156
<211> 20
<212> RNA
<213> Human

<400> 156
gcucuaauac cuugagagca 20

<210> 157
<211> 22
<212> RNA
<213> Human

<400> 157
cuuugagacc ucaaauccug uu 22

<210> 158
<211> 25
<212> RNA
<213> Human

<400> 158
cuuuauuuuc cuuucaucuc ugggc 25

<210> 159
<211> 27
<212> RNA
<213> Human

<400> 159
aucguuucuu cacggacagu gugcugg 27

<210> 160
<211> 24
<212> RNA
<213> Human

<400> 160
gggcuuguga gacaugagug auuu 24

<210> 161
<211> 22
<212> RNA
<213> Human

<400> 161
accuucagag gacuccucuu gc 22

<210> 162
<211> 25
<212> RNA
<213> Human

<400> 162
uauguguuac cuacccuugu cgguc 25

<210> 163
<211> 20
<212> RNA
<213> Human

<400> 163
ggagagagcu uccuguagcu 20

<210> 164
<211> 23
<212> RNA
<213> Human

<400> 164
ucacccuuuc cacaggcguu gca 23

<210> 165
<211> 20
<212> RNA
<213> Human

<400> 165
uuugugucuu ucugagaaac 20

<210> 166
<211> 20
<212> RNA
<213> Human

<400> 166
aaagacuuac cuuaagauac 20

<210> 167
<211> 20
<212> RNA
<213> Human

<400> 167
aucugucaaa ucgccugcag 20

<210> 168
<211> 20
<212> RNA
<213> Human

<400> 168
uuaccuugac uugcucaagc 20

<210> 169
<211> 20
<212> MRNA
<213> Human

<400> 169
uccagguuca agugggauac 20

<210> 170
<211> 25

<212> RNA
<213> Human

<400> 170
gcucuucugg gcuuauggga gcacu 25

<210> 171
<211> 27
<212> RNA
<213> Human

<400> 171
accuuuaucc acuggagauu ugucugc 27

<210> 172
<211> 21
<212> RNA
<213> Human

<400> 172
uuccaccagu aacugaaaca g 21

<210> 173
<211> 29
<212> RNA
<213> Human

<400> 173
ccacucagag cucagaucuu cuaacuucc 29

<210> 174
<211> 27
<212> RNA
<213> Human

<400> 174
cuucaacuca gagcucagau cuucuaa 27

<210> 175
<211> 25
<212> RNA
<213> Human

<400> 175
gggauccagu auacuuacag gcucc 25

<210> 176
<211> 26
<212> RNA
<213> Human

<400> 176
accagaguaa cagucugagu aggagc 26

<210> 177
<211> 23
<212> DNA
<213> Human

<400> 177
cucauaccuu cugcuugaug auc 23

<210> 178
<211> 24
<212> RNA
<213> Human

<400> 178
uucuguccaa gcccgguuga aauc 24

<210> 179
<211> 25
<212> RNA
<213> Human

<400> 179
acaucaagga ggauggcauu ucuag 25

<210> 180
<211> 25
<212> RNA
<213> Human

<400> 196
acaucaagga agauggcauu ucuag 25

<210> 181
<211> 30
<212> RNA
<213> Human

<400> 181
cuccaacauc aaggaagaug gcauuucuag 30

<210> 182
<211> 25
<212> RNA
<213> Human

<400> 182
aucauuuuuu cucauaccuu cugcu 25

<210> 183
<211> 36
<212> DNA
<213> Human

<400> 183
aucauuuuuu cucauaccuu cugcuaggag cuaaaa 36

<210> 184
<211> 21
<212> RNA
<213> Human

<400> 184
cacccaccau cacccucugu g 21

<210> 185
<211> 22
<212> RNA
<213> Human

<400> 185
aucaucucgu ugauauccue aa 22

<210> 186
<211> 21
<212> RNA
<213> Human

<400> 186
uccugcauug uugccuguaa g 21

<210> 187
<211> 30
<212> RNA
<213> Human

<400> 187
uccaacuggg gacgccucug uuccaaaucc 30

<210> 188
<211> 21
<212> RNA
<213> Human

<400> 188
acuggggacg ccucuguucc a 21

<210> 189
<211> 20
<212> RNA
<213> Human

<400> 189
ccguaaugau uguucuagcc 20

<210> 190
<211> 20
<212> RNA
<213> Human

<400> 190
uguuaaaaaa cuuacuucga 20

<210> 191
<211> 25
<212> RNA
<213> Human

<400> 191
cauucaacug uugccuccgg uucug 25

<210> 192
<211> 24

<212> RNA
<213> Human

<400> 192
cuguugccuc cgguucugaa ggug 24

<210> 193
<211> 31
<212> RNA
<213> Human

<400> 193
cauucaacug uugccuccgg uucugaaggu g 31

<210> 194
<211> 21
<212> RNA
<213> Human

<400> 194
uacuaaccuu gguuucugug a 21

<210> 195
<211> 25
<212> RNA
<213> Human

<400> 195
cugaaggugu ucuuguacuu caucc 25

<210> 196
<211> 27
<212> RNA
<213> Human

<400> 196
uguauaggga cccuccuucc augacuc 27

<210> 197
<211> 25
<212> RNA
<213> Human

<400> 197
cuaaccuugg uuucugugau uuucu 25

<210> 198
<211> 27
<212> RNA
<213> Human

<400> 198
gguaucuuug auacuaaccu ugguuuc 27

<210> 199
<211> 22
<212> RNA
<213> Human

<400> 199
auucuuucaa cuagaauaaa ag 22

<210> 200
<211> 25
<212> RNA
<213> Human

<400> 200
gauucugaau ucuuucaacu agaau 25

<210> 201
<211> 20
<212> RNA
<213> Human

<400> 201
aucccacuga uucugaauuc 20

<210> 202
<211> 22
<212> RNA
<213> Human

<400> 202
uuggcucugg ccuguccuaa ga 22

<210> 203
<211> 29
<212> RNA
<213> Human

<400> 203
cucuuuuccag guucaagugg gauacuagc 30

<210> 204
<211> 31
<212> RNA
<213> Human

<400> 204
caagcuuuuc uuuuaguugc ugcucuuuuc c 31

<210> 205
<211> 30
<212> RNA
<213> Human

<400> 205
uauucuuuug uucuucuagc cuggagaaag 30

<210> 206
<211> 28
<212> RNA
<213> Human

<400> 206
cugcuuccuc caaccauaaa acaaauuc 28

<210> 207
<211> 26
<212> RNA
<213> Human

<400> 207
ccaaugccau ccuggaguuc cuguaa 26

<210> 208
<211> 20
<212> RNA
<213> Human

<400> 208
uccuguagaa uacuggcauc 20

<210> 209
<211> 27
<212> RNA
<213> Human

<400> 209
ugcagaccuc cugccaccgc agauuca 27

<210> 210
<211> 20
<212> RNA
<213> Human

<400> 210
cuaccucuuu uuucugucug 20

<210> 211
<211> 20
<212> RNA
<213> Human

<400> 211
uguuuuugag gauugcugaa 20

<210> 212
<211> 84
<212> MRNA
<213> Human

<400> 212

cagcaguagu ugucaucugc ucaacuggca gaauucgauc caccggcugu ucaagccuga 60

gcugaucugc ucgcaucuug cagu 84

Claims

1. An isolated antisense oligonucleotide that binds to human dystrophin pre-mRNA wherein said oligonucleotide is 20 to 31 nucleotides in length and is an oligonucleotide that is specifically hybridizable to an exon 53 target region of the Dystrophin gene designated as annealing site H53A (+23+47), annealing site H53A (+39+69), or both,

wherein said antisense oligonucleotide is a morpholino antisense oligonucleotide, and,
wherein said oligonucleotide induces exon 53 skipping.

2. An antisense oligonucleotide according to claim 1 selected from SEQ ID NOS: 192, 193 and 195, optionally wherein the uracil bases (U) are thymine bases (T).

3. The antisense oligonucleotide of claim 1 or 2, wherein the oligonucleotide is chemically linked to one or more moieties or conjugates that enhance the activity, cellular distribution, or cellular uptake of the antisense oligonucleotide.

4. The antisense oligonucleotide of claim 3, wherein the oligonucleotide is chemically linked to a polyethylene glycol chain.

5. The antisense oligonucleotide of claim 1 or 2 comprising a 5-substituted pyrimidine base.

6. The antisense oligonucleotide of claim 1 or 2 comprising a 5-methylcytosine base.

7. A composition, comprising an antisense oligonucleotide of any one of claims 1-6 and a saline solution that includes a phosphate buffer.

8. An antisense oligonucleotide of any one of claims 1-6, or a composition of claim 7, for use in a method of treatment of muscular dystrophy.

9. The antisense oligonucleotide or composition for use according to claim 8, wherein the muscular dystrophy is Duchenne Muscular Dystrophy.

Patentansprüche

1. Isoliertes Antisense-Oligonucleotid, das an menschliche Dystrophin-Prä-mRNA bindet, worin das Oligonucleotid 20 bis 31 Nucleotide lang ist und ein Oligonucleotid ist, das spezifisch an eine Exon-53-Target-Region des Dystrophin-Gens hybridisierbar ist, die als Annealing-Stelle H53A (+23+47), Annealing-Stelle H53A (+39+69) oder beides bezeichnet wird,
worin das Antisense-Oligonucleotid ein Morpholino-Antisense-Oligonucleotid ist und
worin das Oligonucleotid ein Überspringen von Exon 53 induziert.

2. Antisense-Oligonucleotid nach Anspruch 1, ausgewählt aus den SEQ-ID Nr. 192, 193 und 195, worin die Uracil-Basen (U) gegebenenfalls Thymin-Basen (T) sind.

3. Antisense-Oligonucleotid nach Anspruch 1 oder 2, worin das Oligonucleotid chemisch an eine oder mehrere Gruppierungen oder Konjugate gebunden ist, die die Aktivität, Zellverteilung oder Zellaufnahme des Antisense-Oligonucleotids verbessern.

4. Antisense-Oligonucleotid nach Anspruch 3, worin das Oligonucleotid chemisch an eine Polyethylenglykol-Kette gebunden ist.

5. Antisense-Oligonucleotid nach Anspruch 1 oder 2, umfassend eine 5-substituierte Pyrimidin-Base.

6. Antisense-Oligonucleotid nach Anspruch 1 oder 2, umfassend eine 5-Methylcytosin-Base.

7. Zusammensetzung, umfassend ein Antisense-Oligonucleotid nach einem der Ansprüche 1-6 und eine Kochsalzlösung, die einen Phosphatpuffer umfasst.

8. Antisense-Oligonucleotid nach einem der Ansprüche 1-6 oder Zusammensetzung nach Anspruch 7 zur Verwendung in einem Verfahren zur Behandlung von Muskeldystrophie.

9. Antisense-Oligonucleotid oder Zusammensetzung zur Verwendung nach Anspruch 8, worin die Muskeldystrophie eine Muskeldystrophie Duchenne ist.

Revendications

1. Oligonucléotide antisens isolé qui se lie à un pré-ARNm de la dystrophine humaine, où ledit oligonucléotide possède une longueur de 20 à 31 nucléotides et est un oligonucléotide qui peut spécifiquement s'hybrider à une région cible de l'exon 53 du gène de la dystrophine désignée en tant que site d'annelage H53A (+23+47), site d'annelage H53A (+39+69), ou les deux,
où ledit oligonucléotide antisens est un oligonucléotide antisens morpholino, et,
où ledit oligonucléotide induit un saut de l'exon 53.

2. Oligonucléotide antisens selon la revendication 1 sélectionné parmi SEQ ID NO: 192, 193 et 195, facultativement où les bases uracile (U) sont des bases thymine (T).

3. Oligonucléotide antisens selon la revendication 1 ou 2, où l'oligonucléotide est chimiquement lié à un ou plusieurs fragments ou conjugués qui améliorent l'activité, la distribution cellulaire, ou l'absorption cellulaire de l'oligonucléotide antisens.

4. Oligonucléotide antisens selon la revendication 3, où l'oligonucléotide est chimiquement lié à une chaîne polyéthylène glycol.

5. Oligonucléotide antisens selon la revendication 1 ou 2, comprenant une base pyrimidine 5-substituée.

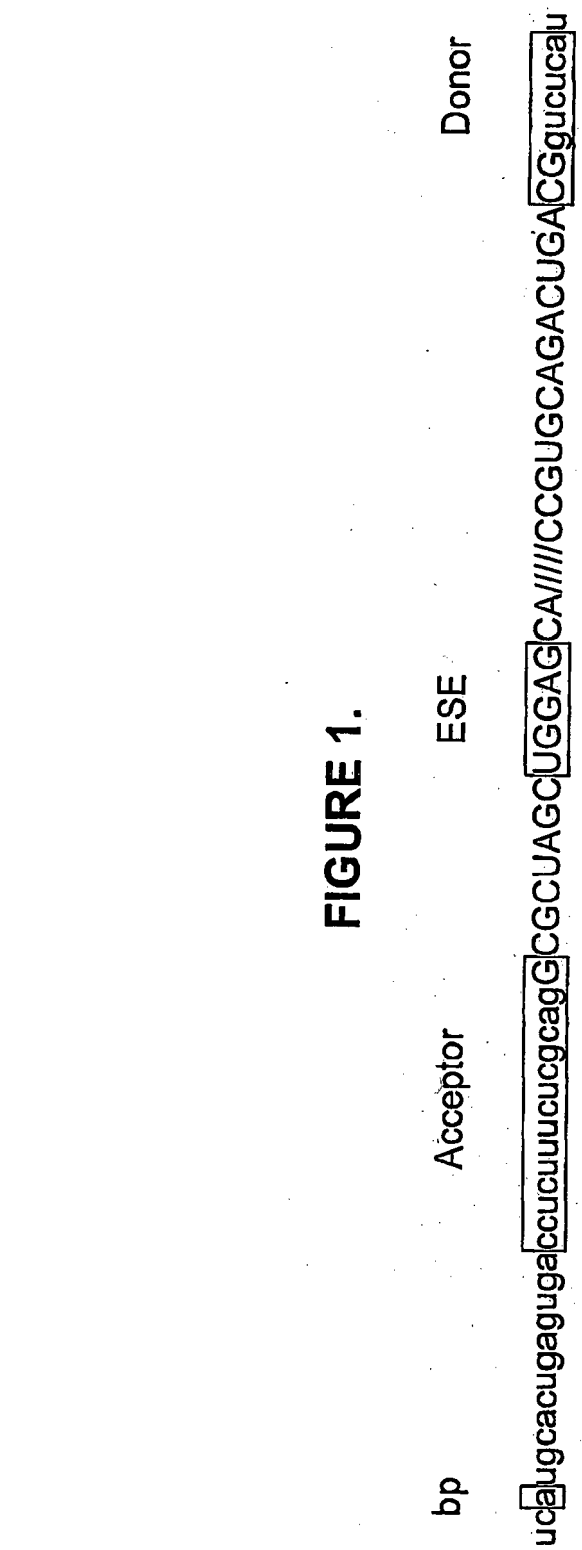
6. Oligonucléotide antisens selon la revendication 1 ou 2, comprenant une base 5-méthylcytosine.

7. Composition, comprenant un oligonucléotide antisens selon l'une quelconque des revendications 1-6 et une solution saline qui comprend un tampon phosphate.

8. Oligonucléotide antisens selon l'une quelconque des revendications 1-6, ou composition selon la revendication 7, destiné(e) à être utilisé(e) dans un procédé de traitement de la dystrophie musculaire.

9. Oligonucléotide antisens ou composition destiné(e) à être utilisé(e) selon la revendication 8, où la dystrophie musculaire est la dystrophie musculaire de Duchenne.

FIGURE 1.


bp
Acceptor
ESE
Donor
ucaugcacugagaugagccucuuucucgcagGCGCUAGCUGGAGGCA////CCGUGCAGACUGACGgucucau

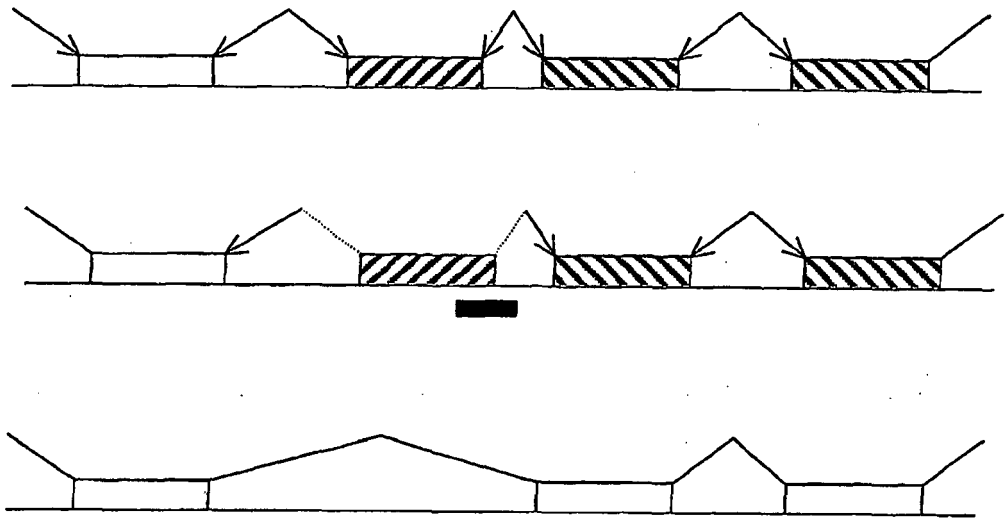
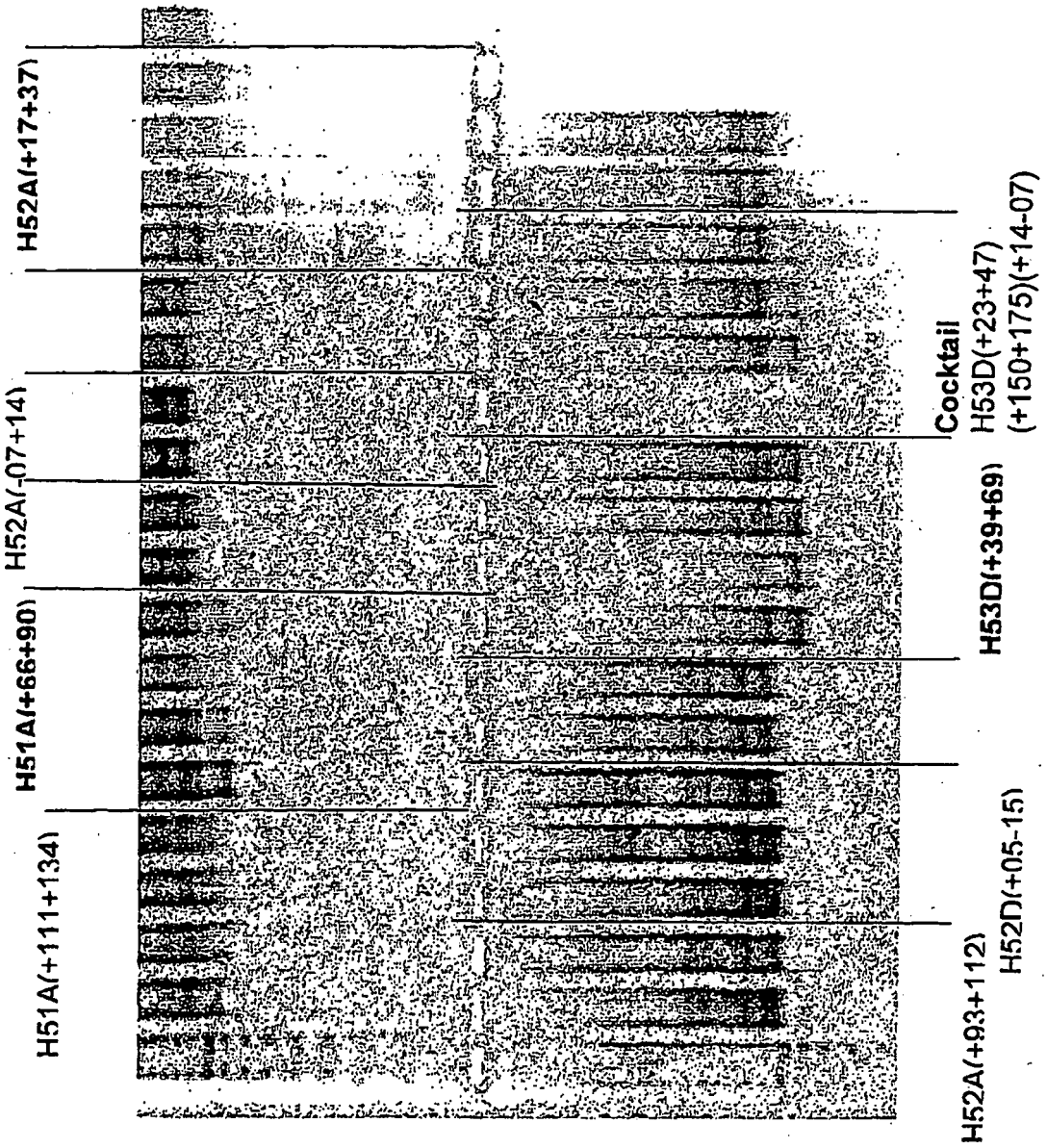


FIGURE 2

FIGURE 3


H52A(+17+37)
H52A(-07+14)
H51A(+66+90)
H51A(+111+134)
Cocktail
H53D(+23+47)
(+150+175)(+14-07)
H53D(+39+69)
H52D(+05-15)
H52A(+93+112)

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- CA 2507125 **[0016]**
- WO 2004083446 A **[0016]**
- US 5149797 A **[0059]**
- US 4458066 A **[0066]**
- US 20040248833 A **[0075]**
- US 6806084 B **[0077]**
- EP 10004274 A **[0107]**
- EP 05754344 A **[0107]**
- AU 2005000943 W **[0107]**
- AU 20040903474 **[0107]**

Non-patent literature cited in the description

- **SIERAKOWSKA H et al.** *Proc Natl Acad Sci USA,* 1996, vol. 93, 12840-12844 **[0004]**
- **WILTON SD et al.** *Neuromusc Disorders,* 1999, vol. 9, 330-338 **[0004]**
- **VAN DEUTEKOM JC et al.** *Human Mol Genet,* 2001, vol. 10, 1547-1554 **[0004]**
- **SHERRAT TG et al.** *Am J Hum Genet,* 1993, vol. 53, 1007-1015 **[0006]**
- **LU QL et al.** *Nature Medicine,* 2003, vol. 9, 1009-1014 **[0006]**
- **AARTSMA-RUS A et al.** *Am J Hum Genet,* 2004, vol. 74, 83-92 **[0006]**
- **MATSUO et al.** *J Clin Invest.,* 1991, vol. 87, 2127-2131 **[0009]**
- **TAKESHIMA et al.** *J. Clin. Invest.,* 1995, vol. 95, 515-520 **[0009]**
- **DUNCKLEY et al.** *Nucleosides & Nucleotides,* 1997, vol. 16, 1665-1668 **[0010]**
- **DUNCKLEY et al.** *Human Mol. Genetics,* 1998, vol. 5, 1083-90 **[0011]**
- **ERRINGTON et al.** *J Gen Med,* 2003, vol. 5, 518-527 **[0012]**
- **WILTON et al.** *Neuromuscular Disorders,* 1999, vol. 9, 330-338 **[0014]**
- **MANN CJ et al.** *J Gen Med,* 2002, vol. 4, 644-654 **[0015]**
- **AARTSMA-RUS et al.** *Human Molecular Genetics,* 2003, vol. 12, 907-14 **[0016]**
- **AARTSMA-RUS et al.** *Neuromuscular Disorders,* 2002, vol. 12, 71-7 **[0016]**
- **MANN et al.** *J Gen Med,* 2002, vol. 4, 644-654 **[0034]**
- Improved antisense oligonucleotide induced exon skipping in the mdx mouse model of muscular dystrophy. *J Gen Med,* vol. 4, 644-654 **[0046]**
- **BEAUCAGE et al.** *Tetrahedron Letters,* 1981, vol. 22, 1859-1862 **[0067]**
- **MARTIN.** Remington's Pharmaceutical Sciences. Mack Publishing Co, 1990 **[0071]**
- **MARTIN.** Remington's Pharmaceutical Sciences. Mack Publishing Co, 1990, 1435-1712 **[0072]**
- **MANN CJ et al.** Antisense-induced exon skipping and the synthesis of dystrophin in the mdx mouse. *Proc., Natl. Acad. Science,* 2001, vol. 98 (1), 42-47 **[0076]**
- **GEBSKI et al.** *Human Molecular Genetics,* 2003, vol. 12 (15), 1801-1811 **[0076]**
- **FRALEY et al.** *Trends Biochem. Sci.,* 1981, vol. 6, 77 **[0079]**
- **MANNINO et al.** *Biotechniques,* 1988, vol. 6, 682 **[0080]**
- **FRIEDMANN.** *Science,* 1989, vol. 244, 1275-1280 **[0083]**
- **ROSENBERG.** *Cancer Research,* 1991, vol. 51 (18), 5074S-5079S **[0084]**
- **ROSENFELD et al.** *Cell,* 1991, vol. 68, 143-155 **[0084]**
- **ROSENFELD et al.** *Science,* 1991, vol. 252, 431-434 **[0084]**
- **BRIGHAM et al.** *Am. J. Med. Sci.,* 1989, vol. 298, 278-281 **[0084]**
- **NABEL et al.** *Science,* 1990, vol. 249, 1285-1288 **[0084]**
- **HAZINSKI et al.** *Am. J. Resp. Cell Molec. Biol.,* 1991, vol. 4, 206-209 **[0084]**
- **WANG ; HUANG.** *Proc. Natl. Acad. Sci. (USA),* 1987, vol. 84, 7851-7855 **[0084]**
- **WU ; WU.** *J. Biol. Chem.,* 1988, vol. 263, 14621-14624 **[0084]**
- **WOLFF et al.** *Science,* 1990, vol. 247, 1465-1468 **[0084]**
- *Am. J. Med. Sci.,* 1989, vol. 298, 278-281 **[0084]**
- *Clinical Research,* 1991, 39 **[0084]**
- **ANDERSON.** *Science,* 1992, vol. 256, 808-813 **[0084]**
- **SAMBROOK et al.** Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, 1989 **[0093]**
- DNA Cloning: A Practical Approach. MRL Press, Ltd, 1985 **[0093]**

- **AUSUBEL, F. ; BRENT, R. ; KINGSTON, R.E. ; MOORE, D.D. ; SEIDMAN, J.G. ; SMITH, J.A. ; STRUHL, K.** Current Protocols in Molecular Biology. Greene Publishing Associates/Wiley Intersciences, 2002 **[0093]**

EXHIBIT 11

LEDERER & KELLER

Patentanwälte Partnerschaft mbB
European Patent and Trademark Attorneys

DR. A. VAN DER WERTH
(1934 – 1974)

DR. FRANZ LEDERER
(1963 – 2012)

DR. GÜNTER KELLER
Dipl.-Biol.

DR. MICHAEL BEST
Dipl.-Chem.

DR. STEPHAN TEIPEL
Dipl.-Chem.

DR. GEORG KALHAMMER
Dipl.-Biochem.

80538 MÜNCHEN
Unsöldstraße 2
Telefon +49 89 21 23 99 0
Telefax +49 89 21 23 99 22
E-Mail info@lederer-keller.de

European Patent Office

80298 Munich

Your reference	Our reference	25 August 2016
	ESP00349NIP	K/sh

Re.: European Patent No. 2 206 781 B1
European Patent Application No. 10004274.6
Patentee: The University of Western Australia
Opponent: Nippon Shinyaku Co., Ltd.

On behalf of

Nippon Shinyaku Co. Ltd.
14, Nishinosho-Monguchi-cho, Kisshoin,
Minami-ku, Kyoto-shi,
Kyoto 601-8550
Japan

OPPOSITION

is lodged according to Article 99 EPC against the above-referenced patent titled

"Antisense oligonucleotides for inducing exon skipping and methods for use thereof".

The opposition fee amounting to EUR 785,-- is to be debited from our deposit account no. 28000381. A corresponding payment order is enclosed.

The opponent has appointed us as his representatives and it is requested to effect all notifications to our address.

The European patent EP 2 206 781 is opposed in its full extent (claims 1 to 9).

The opposition is based on the grounds of Article 100(a), 100(b) and 100(c) EPC. In particular it is submitted that the patent lacks inventive step. It does not disclose the invention in a manner sufficiently clear and complete for it to be carried out by the skilled person and its subject matter extends beyond the content of the application as originally filed.

It is requested to revoke the European patent in total. Oral proceedings in accordance with Article 116 EPC are requested in the event that the Opposition Division does not reach the decision to revoke the patent on the basis of the written submission of the opponent.

Detailed statement of the grounds for opposition

I. The cited prior art

For substantiation of the opposition it is referred to the prior art documents cited in the opposed patent and during the granting proceedings and the following documents. In particular it is referred to the following documents:

D1 Corey et al., Genome Biology, 2001, 2(5) 1015.1-1015.3

D2 AU 2004903474 (priority document)

D3 WO 2004/083432

D4 WO 2004/048570 (≙ EP 1 568 769)

D5 CA 2 507 125

D6 Aartsma-Rus et al., Human Molecular Genetics (2003), vol. 12, no. 8; pp 907-914

D7 Aartsma-Rus et al., Neuromuscular Disorders vol.12,S71-S77(2002)

D8 experimental report

II. The subject matter of EP 2 206 781

Claim 1 of the opposed patent relates to

a) an isolated antisense oligonucleotide that
   - a_1) binds to human dystrophin pre-mRNA,
   - a_2) wherein said oligonucleotide is 20 to 31 nucleotides in length and

b) is an oligonucleotide that is specifically hybridizable to
   - b_1) an exon 53 target region of the Dystrophin gene designated as annealing site H53A (+23+47),
   - b_2) annealing site H53A (+39+69), or
   - b_3) both

c) wherein said antisense oligonucleotide is a morpholino antisense oligonucleotide and

d) wherein said oligonucleotide induces exon 53 skipping.

Claim 2 relates to an oligonucleotide according to claim 1 which is selected from SEQ ID NOs: 192, 193 and 195, optionally wherein the uracil bases (U) are thymine bases (T).

Claim 3 is referred back to claims 1 or 2 and defines the oligonucleotide further in that it is chemically linked to one or more moieties or conjugates that enhance the activity, cellular distribution, or cellular uptake of the antisense oligonucleotide.

Claim 4 is referred back to claim 3 and defines the moieties or conjugates wherein the oligonucleotide is chemically linked to a polyethylene glycol chain.

Claim 5 is referred back to claim 1 or 2 and mentions that it comprises a 5-substituted pyrimidine base.

Claim 6 is referred back to claim 1 or 2 and mentions that it comprises a 5-methylcytosine base.

Claim 7 relates to a composition comprising an oligonucleotide as defined in any of claims 1-6 and a saline solution that includes a phosphate buffer.

Claim 8 relates to an antisense oligonucleotide according to any of claims 1-6, or a composition of claim 7, for use in a method of treatment of muscular dystrophy.

Claim 9 is referred back to claim 8 and defines the muscular dystrophy further in that it is Duchenne Muscular Dystrophy.

It is alleged that the opposed patent provides a target region within exon 53 spanning the annealing sites H53A (+23+47) and/or H53A (+39+69) with a morpholino antisense oligonucleotide ("PMO") of 20 to 31 nucleotides in length. This target region is efficient in vivo splicing (compare with patentee's submission dated September 17, 2014). It is alleged that the key region of exon 53 can be targeted to provide therapeutically effective exon skipping.

III. Article 100(c)/123(2) EPC / added matter

3.1 Intermediate generalization

The granted claims have to be revoked in view of an inadmissible intermediate generalization since by the amendments features were taken out of the initial context and combined with others. Such "intermediate generalization" is considered to be inadmissible (see e.g. T962/98, T1408/04, T461/05 or T1118/10 to name only a few of the relevant decisions).

3.1.1 Three annealing sites

Claim 1 contains three annealing sites, namely

1) H53A (+23+47) which corresponds to feature b_1) and
2) annealing site H53A (+39+69) as annealing site b_2) and furthermore
3) both annealing sites (feature b_3)).

First of all it should be mentioned that feature b_3), namely both annealing sites, is obviously nowhere disclosed. To be more precise, the word "both" in combination with the two annealing sites representing features b_1) and b_2) of claim 1 can nowhere be found in the specification as originally filed (WO 2006/000057).

For the interpretation what the term "both" should mean according to the interpretation of patentee we refer to the submission of patentee's representatives dated September 17, 2014, page 2, first paragraph. This paragraph reads as follows:

"*The solution provided by the present invention is to target a region within exon 53 spanning the annealing sites H53A (+23+47) and/or H53A (+39+69) with a morpholino antisense oligonucleotide ('PMO') of 20-31 nucleotides in length.*"

From this statement it can be concluded that it is the intention of patentee to claim not only one of the two annealing sites (feature b_1) or b_2), respectively), but also the area from nucleotide +23 until +69. This is supported by Figure A which was also submitted on November 17, 2014 by patentee's representatives. This "area" is nowhere disclosed in the application as originally filed. Therefore, Article 123(2) EPC is violated.

According to feature a_2) the claimed oligonucleotide should be 20-31 nucleotides in length. Concerning the length of the oligonucleotides a paragraph on page 21, lines 9-17 of WO 2006/000057 can be found. This passage generally says that the antisense oligonucleotides can be as short as 12 bases whereas such length are not as efficient as longer (20-31 bases) oligonucleotides. The selection of 20-31 bases is therefore a selection of a list of various lengths of oligonucleotides. When we consider the relevant passage on page 62 of WO 2006/000057 there is no range of oligonucleotides provided. Table 39 summarizes 12 antisense oligonucleotides having different lengths. It should be noted that only two target regions are features of the claims, namely b_1) and b_2). The antisense oligonucleotide H53A (+23+47) has a length of 25 bases and the oligonucleotide H53A (+39+69) has a length of 31 bases. Those two embodiments corresponding to SEQ ID NO: 193 and 195 cannot form a basis for the range of 20-31 nucleotides.

It seems that the feature that the antisense oligonucleotide is a morpholino antisense oligonucleotide (feature c) of claim 1) has a potential basis on page 17, line 3 of the published international application. This passage relates, however, to Table 1A wherein all sequences designated as SEQ ID NOs: 1-211 are summarized. Alternatively in the relevant sentence on page 17, lines 2-4, other antisense chemistries such as peptide nucleic acids are, however, mentioned.

A basis for feature d), namely that the oligonucleotide induces exon 53 skipping might potentially be found on page 62 and in particular Table 39 of WO 2006/000057.

In Table 39 there are mentioned several antisense oligonucleotides whereby nine oligonucleotides are designated as H53A. The splice sites at the beginning and end of the exon are, however, different.

Patentee has selected from Table 39, which can be considered as a list, among nine individual oligonucleotides only two and the reason for selecting features b_1) and b_2), respectively, is not evident. When we look at the biological activity (ability to induce skipping) H53A (+39+69) corresponding to SEQ ID NO: 193 (feature b_2)) is said to induce strong skipping to 50 nM. The annealing site of feature b_1), namely H53A (+23+47) corresponding to SEQ ID NO: 195 is said to induce only very faint skipping to 50 nM.

According to claim 2 SEQ ID NO: 192 which corresponds to H53A (+36+62) should also fall under claim 1 and shows only a faint skipping at 50 nM (compare with Table 39).

Considering the other biological activities, namely the ability to induce skipping of other antisense oligonucleotides we noticed that for example H53A (+45+69) can induce faint skipping at 50 nM or H53A (+150+176) can also induce very faint skipping at 50 nM.

Therefore, features b_1) and b_2), respectively, were arbitrarily selected and it seems that the selected oligonucleotides do not fulfill the feature d) of claim 1, namely that the oligonucleotide induces exon 53 skipping since Table 39 shows varying degrees of biological activity down to very faint skipping.

3.1.2 Specifically hybridizable

When considering the term "specifically hybridizable to" in claim 1 it is more likely that also oligonucleotides that have little or faint skipping activities were used as basis. The description only provides a vague definition for "specifically hybridizable" in paragraphs [0052] and [0053], reading, e.g. *"a sufficient degree of complementarity or precise pairing such that stable and specific binding occurs between the oligonucleotide and the DNA or RNA target"* (see para. [0052]).

The specification as originally filed defines on page 23, lines 16-31 the term "specifically hybridizable". An antisense molecule is considered to be specifically hybridizable when binding of this compound to the target DNA or RNA molecule interferes with the normal function of the target DNA or RNA to cause a loss of utility and there is a specific degree of complementarity to avoid non-specific binding of the antisense compound to non-target sequences under conditions in which specific binding is desired, i.e. under physiological conditions in the case of in vivo assays or in the case of in vitro assays.

An oligonucleotide is therefore only "specifically hybridizable" when two conditions are fulfilled:

1) The antisense molecule must specifically hybridize with the target DNA or RNA and

2) Non-specific binding of the antisense compound to non-target sequences must be avoided. It is evident that this second condition requires that it had to be tested whether an oligonucleotide does bind or does not bind in a non-specific way to non-target sequences.

As potential basis for claim 1 pages 62-63 of the originally filed documents may serve. This section can, however, not serve as a basis for the feature "specifically hybridizing" for three reasons:

1) The strength of binding by hybridization and skipping is not the same;

2) Only the antisense oligonucleotide designated as H53A(+39+69) is said to induce strong skipping whereas the other oligonucleotides show only faint or very faint skipping speaking for a comparatively weak hybridization;

3) There is no disclosure whatsoever that those oligonucleotides do not specifically bind to non-target sequences.

Therefore, the passage in the general part of the specification on page 23 cannot be combined with the disclosure of Table 39. With other words: There is no disclosure that the antisense oligonucleotides disclosed on page 62 have the feature of "specifically hybridize to".

3.2 Selection from at least two lists

The opposed patent is to be revoked since granted claim 1 is an inadmissible selection from more than two lists of features which violates the requirements of Article 123(2) EPC (e.g. T223/11, T1651/11).

The combination of features created by patentee is not admissible under established jurisprudence of the Boards of Appeal considering the aspect that the features are selected from several lists. This is not allowable in view of the quoted decisions of the Board of Appeal. Of course there are many more T-decisions which could also be quoted in order to support this argument.

Furthermore, dependent sub-claims 2-9 are directly or indirectly referred back to claim 1. There is no disclosure wherein the features are originally disclosed in combination. Therefore, patentee has selected the features from different lists (e.g. former subclaims) which is a clear violation of the established case law of the Boards of Appeal.

The objected claims have therefore to be revoked, because the presently claimed combination of this feature is not disclosed and the claims are considered to be an artificially created new embodiment (e.g. T 2496/10, point 4.6).

IV. Article 100(b) EPC

Claim 1 relates to oligonucleotides with 20 to 31 nucleotides in length. It is submitted that the claimed invention is not enabled over the whole scope of the claim, in particular for oligonucleotides having 20 to 24 oligonucleotides.

The whole scope of the claim is not supported by the description, in view of the low predictability in the art, especially for the annealing range of +39+69, because the description provides data for only 191(+45+69),192(+39+62) and 193(+39+69) but no others.

These relevant oligomers from Table 39 have the length of 25, 24 and 31 base pairs respectively. In view of this, the description of the opposed patent fails to support whether oligonucleotides having the length of less than 24 base pairs actually have skipping activity. In this context, Corey and Abrams (Genome Biology 2001, 2 (5); 1015.1-1015.3, submitted as document **D1**) teach for morpholino antisense oligonucleotides that "this binding is no tighter than binding of analogous DNA and RNA oligomers, necessitating the use of relatively long 25-base morpholinos for antisense gene inhibition" (page 1015.1, left column, 4th to 2nd lines from the bottom).

In the absence of experimental data for oligonucleotides having the length of less than 24 base pairs, and in view of the teaching by Corey and Abrams (**D1**), it must be concluded that oligonucleotides having the length of less than 24 base pairs are not supported in the description.

Furthermore, it should be stressed that the only oligonucleotide which shows according to Table 39 a strong skipping ability has a length of 31 nucleotides. Contrary thereto shorter oligonucleotides like H53D(+14-07) having a length of 21 oligonucleotides or H53A(-12+10) having a length of 22 oligonucleotides or H53A(+07+26) having a length of 20 oligonucleotides

show only a very faint skipping or no skipping ability at all. Therefore, the opposed patent is to be revoked in view of lack of enabling disclosure.

V. Priority

The opposed patent claims the priority of the Australian patent application no. 2004903474 as filed on June 28, 2004. None of the claims is entitled to the priority of the Australian patent application. Claim 1 relates to two annealing sites (feature b_1) and feature b_2)) or both (feature b_3)). None of those annealing sites has a basis in the priority document.

The oligomers disclosed in the priority application are summarized in Table 1 in the specification of the priority document submitted as **D1**. Table 1 is shown in the following:

Brief Description of the Sequence listings

SEQ ID	SEQUENCE	NUCLEOTIDE SEQUENCE (5' - 3')
1	H8A(-06+18)	GAU AGG UGG UAU CAA CAU CUG UAA
2	H8A (-03+18)	GAU AGG UGG UAU CAA CAU CUG
3	H8A(-07+18)	GAU AGG UGG UAU CAA CAU CUG UAA G
4	H8A(-06+14)	GGU GGU AUC AAC AUC UGU AA
5	H8A(-10+10)	GUA UCA ACA UCU GUA AGC AC
6	H7A(+45+67)	UGC AUG UUC CAG UCG UUG UGU GG
7	H7A(+02+26)	CAC UAU UCC AGU CAA AUA GGU CUG G
8	H7D(+15-10)	AUU UAC CAA CCU UCA GGA UCG AGU A
9	H7A(-18+03)	GGC CUA AAA CAC AUA CAC AUA
10	C6A(-10+10)	CAU UUU UGA CCU ACA UGU GG
11	C6A(-14+06)	UUU GAC CUA CAU GUG GAA AG
12	C6A(-14+12)	UAC AUU UUU GAC CUA CAU GUG GAA AG

13	C6A(-13+09)	AUU UUU GAC CUA CAU GGG AAA G
14	CH6A(+69+91)	UAC GAG UUG AUU GUC GGA CCC AG
15	C6D(+12-13)	GUG GUC UCC UUA CCU AUG ACU GUG G
16	C6D(+06-11)	GGU CUC CUU ACC UAU GA
17	H6D(+04-21)	UGU CUC AGU AAU CUU CUU ACC UAU
18	H6D(+18-04)	UCU UAC CUA UGA CUA UGG AUG AGA
19	H4A(+13+32)	GCA UGA ACU CUU GUG GAU CC
20	H4D(+04-16)	CCA GGG UAC UAC UUA CAU UA
21	H4D(-24-44)	AUC GUG UGU CAC AGC AUC CAG

Table 1: Description of 2'-O-methyl phosphorothioate antisense oligonucleotides that have been used to date to study induced exon skipping during the processing of the dystrophin pre-mRNA.

Only skipping of exons 8, 7, 6 and 4 is attempted in the priority application.

Thus, exon 53 skipping and oligomers therefor, H53A (+23+47) and H53A (+39+69), are not disclosed in the priority application and the claims are therefore not entitled to the priority of June 28, 2005.

As the H53A (+23+47) and H53A (+39+69), as exemplary embodiments of claim 1, are not entitled to the priority, inventiveness of claims 1 to 9 is to be evaluated as of the international filing date of **28 June 2005**.

VI. Inventive Step

6.1 WO 2004/083432 (**D3**)

Claim 1 of the opposed patent relates to

a) an isolated antisense oligonucleotide that
   - a_1) binds to human dystrophin pre-mRNA,
   - a_2) wherein said oligonucleotide is 20 to 31 nucleotides in length and

b) is an oligonucleotide that is specifically hybridizable to
   - b_1) an exon 53 target region of the Dystrophin gene designated as annealing site H53A (+23+47),

b_2) annealing site H53A (+39+69), or

b_3) both

c) wherein said antisense oligonucleotide is a morpholino antisense oligonucleotide and

d) wherein said oligonucleotide induces exon 53 skipping.

D3 (WO 2004/083432) discloses an antisense oligomer named "h53AON1", which enables exon 53 skipping and has the following sequence (see Table 2 on page 48):

Table 2
Characteristics of the AONs used to study the targeted skipping of 15 different DMD exons*

Name	Antisense sequence (5'-3')	Length (bp)	G/C%	U/C%	Exon skip	Transcript
h53AON 1	cuguugccuccgguucug	18	61	72	+	OF

The complementary sequence of h53AON1 is CAGAACCGGAGGCAACAG. This sequence is completely encompassed in the claimed annealing sites b_2) and b_3) of the subject patent (see below where the yellow highlighted part of b_2) corresponds to the complementary sequence of h53AON1).

b_2): **CACCTTCAGAACCGGAGGCAACAGTTGAATG**

Therefore, **D3** discloses antisense oligonucleotide having feature b).

In addition, h53AON1 is shown to have a high skipping activity (see Figure 1E etc.). **D3** also discloses that the oligomers can be modified with morpholino phosphorodiamidate (see page 10, line 11 to page 11, line 4). Thus, the features c) and d), i.e. "wherein said antisense oligonucleotide is a morpholino antisense oligonucleotide, and, wherein said oligonucleotide induces exon 53 skipping", are disclosed in **D3**.

h53AON1 per se has a length of 18 oligonucleotides and does not satisfy requirement a_2) "wherein said oligonucleotide is 20 to 31 nucleotides in length".

However, **D3** states "*[c]urrently, many different compounds are available that mimic hybridization characteristics of oligonucleotides. Such a compound is also suitable for the present invention is such equivalent comprises similar hybridization characteristics in kind not necessarily in amount....As mentioned, oligonucleotides of the invention do not have to consist of only oligonucleotides that contributes to hybridization to the targeted exon. There may be additional material and/or nucleotides added.*" (see page 12, lines 11 to 20 of **D3**).

Based on the above statement, an average person of the skill in the art would have easily conceived of a "modification" of h53AON1 having additional nucleotides added, which may well be of the length of 20 to 31 nucleotides. Further, when two nucleotides are added to h53AON1 to make a 20mer equivalent, with any combination of those two nucleotides, at least 90% of the entire stretch of the 20mer equivalent is still completely identical to the counterpart region of the b_2) annealing site. Thus, it is clear that such 20mer is "specifically hybridizable to" b_2).

Considering the teaching by Corey and Abrams (**D1**) wherein a longer oligonucleotide is recommended for skipping (see supra), an average person skill in the art would have been motivated to modify h53AON1 by adding extra nucleotides to make it longer.

Therefore, the subject matter of claim 1 lacks inventive step over **D3** since **D3** contains the suggestion to elongate short nucleotides.

5.2 WO 2004/048570 (**D4**) in view of **D3**

The international patent application WO 2004/048570 was originally published in Japanese language on September 30, 2004. The document is therefore pre-published prior art. The English translation of this document was published as EP 1 568 769 and we refer in the following to this English document as **D4**.

D4 discloses oligomers enabling exon 53 skipping (see Examples 79-87), among which AO95 of Example 87 having SEQ ID NO: 75: corresponds to the 30th to 47th nucleotides of exon 53.

SEQD ID NO:75 of the sequence listing shows the nucleotide sequence of the oligonucleotide prepared in Example 87 (A095) (page 216, line 29).

<210> 75
<211> 18
<212> DNA
<213> synthetic oligonucleotide

<400> 75
ctgaaggtgt tcttgtac

The complementary sequence of AO95 is GTACAAGAACACCTTCAG.

This sequence is completely encompassed in the claimed annealing sites b_1) and b_3) of the subject patent (see below where the yellow highlighted part of b_1) corresponds to the complementary sequence of AO95).

b_1): **GGATGAAGTACAAGAACACCTTCAG**

Thus, AO95 is contained within SEQ195 (+23+47) of the subject claim. AO95 is demonstrated to have a skipping activity (see Figure 19 and para. [0319] of EP1568769A1). **D4** also discloses that the compound may be of morpholine salt (see para. [0046] of EP1568769A1).

AO95 per se is in the length of 18mer and does not comply with requirement a_2) "wherein said oligonucleotide is 20 to 31 nucleotides in length". However, **D3** states "*[c]urrently, many different compounds are available that mimic hybridization characteristics of oligonucleotides. Such a compound is also suitable for the present invention is such equivalent comprises similar hybridization characteristics in kind not necessarily in amount....As mentioned, oligonucleotides of the invention do not have to consist of only oligonucleotides that contributes to hybridization to the targeted exon. There may be additional material and/or nucleotides added.*" (see page 12, lines 11 to 20 of **D3**).

Based on the above statements of **D3**, a person skilled in the art would have easily conceived of a variant of AO95 having additional nucleotides added, which may well be of the length of 20-31. Further, when two nucleotides are added to AO95 to make a 20mer equivalent, with any combination of those two nucleotides, at least 90% of the entire stretch of the 20mer equivalent is still completely identical to the counterpart region of the b_1) annealing site. Thus, it is clear that the 20mer equivalent "specifically hybridizable to" b_1).

A person skilled in the art starting from the specific disclosure as explained above in document **D3** or **D4**, respectively, and looking for an alternative embodiment has a strong motivation to elongate the 18mers as disclosed in **D3** or **D4**, respectively, by at least 2-10 oligonucleotides since it can be expected that the binding of the oligonucleotide will be improved by such an elongation as taught by **D3** or **D1**, respectively.

Therefore, the claimed invention is rendered obvious for the average person skilled in the art by either a combination of **D3** and **D4,** or a combination of **D3** and **D1**.

5.3 The subject matter of claim 1 does not exert superior effect in full scope.

"h53AON1" has been cited as a prior art by the Examining Division of EPO in the examination stage of the subject patent. In the opinion attached to the extended European search report dated January 2, 2013 the searching authority cited inter alia documents D1, D2 and D3. These documents are submitted as document **D5** (CA 2 507 125), **D6** (Human Molecular Genetics, vol. 12 (2003), pages 907-914) and **D7** (Neuromuscular Disorders, vol. 12, January 2002, pages S71-S77). In section 3 the searching authority concluded that all of documents D1-D3 (now corresponding to **D5-D7**) disclose antisense oligonucleotides targeting exon 53 in the region corresponding to H53A (+23+47) and H53A (+39+69) for inducing exon skipping in the dystrophin gene in order to treat DMD.

Regarding "h53AON1", the University of Western Australia argued that H53A (+39+69) is superior over h53AON1 in terms of skipping activity (see Annex 2 - response filed by the representatives of the University of Western Australia dated 5 November 2013).

Nippon Shinyaku (NS) conducted experiments to see whether any oligomers falling within the scope of claim 1 of the subject patent have superior activity over h53AON1 and found that some oligomers have inferior activities to h53AON1. The experimental report was performed by opponent and is submitted as document **D 8**.

The data as shown in the experimental report submitted as **D8**, confirm that the alleged superior activity as argued by the patentee in the examination stage is not obtainable over the whole scope of the claim, and thus the claimed subject matter is not inventive.

Since no superior effect has been shown over the closest prior art the object of the opposed patent is to provide an alternative oligonucleotide. Such an alternative is, however, rendered obvious by the prior art, in particular by a combination of **D3** and **D1** or **D4** and **D3**.

The claimed subject matter is not inventive in view of the relevant prior art.

Conclusion

Since the opposed patent violates the European patent convention in several respect, the complete revocation of the opposed patent is justified.

Dr. G. Keller

Enc.:
1 copy
payment order
D1-D8, 2-fold

EXHIBIT 12

FDA NEWS RELEASE

FDA grants accelerated approval to first drug for Duchenne muscular dystrophy

For Immediate Release:

September 19, 2016

Español (/news-events/comunicados-de-prensa/la-fda-concede-la-aprobacion-acelerada-al-primer-medicamento-para-la-distrofia-muscular-de-duchenne)

The U.S. Food and Drug Administration today approved Exondys 51 (eteplirsen) injection, the first drug approved to treat patients with Duchenne muscular dystrophy (DMD). Exondys 51 is specifically indicated for patients who have a confirmed mutation of the dystrophin gene amenable to exon 51 skipping, which affects about 13 percent of the population with DMD.

“Patients with a particular type of Duchenne muscular dystrophy will now have access to an approved treatment for this rare and devastating disease,” said Janet Woodcock, M.D., director of the FDA’s Center for Drug Evaluation and Research. “In rare diseases, new drug development is especially challenging due to the small numbers of people affected by each disease and the lack of medical understanding of many disorders. Accelerated approval makes this drug available to patients based on initial data, but we eagerly await learning more about the efficacy of this drug through a confirmatory clinical trial that the company must conduct after approval.”

DMD is a rare genetic disorder characterized by progressive muscle deterioration and weakness. It is the most common type of muscular dystrophy (https://www.ninds.nih.gov/Disorders/All-Disorders/Muscular-Dystrophy-Information-Page). DMD is caused by an absence of dystrophin, a protein that helps keep muscle cells intact. The first symptoms are usually seen between three and five years of age, and worsen over time. The disease often occurs in people without a known family history of the condition and primarily affects boys, but in rare cases it can affect girls. DMD occurs in about one out of every 3,600 male infants worldwide.

People with DMD progressively lose the ability to perform activities independently and often require use of a wheelchair by their early teens. As the disease progresses, life-threatening heart and respiratory conditions can occur. Patients typically succumb to the disease in their 20s or 30s; however, disease severity and life expectancy vary.

Exondys 51 was approved under the accelerated approval pathway, which provides for the approval of drugs that treat serious or life-threatening diseases and generally provide a meaningful advantage over existing treatments. Approval under this pathway can be based on adequate and well-controlled studies showing the drug has an effect on a surrogate endpoint that is reasonably likely to predict clinical benefit to patients (how a patient feels or functions or whether they survive). This pathway provides earlier patient access to promising new drugs while the company conducts clinical trials to verify the predicted clinical benefit.

The accelerated approval of Exondys 51 is based on the surrogate endpoint of dystrophin increase in skeletal muscle observed in some Exondys 51-treated patients. The FDA has concluded that the data submitted by the applicant demonstrated an increase in dystrophin production that is reasonably likely to predict clinical benefit in some patients with DMD who have a confirmed mutation of the dystrophin gene amenable to exon 51 skipping. A clinical benefit of Exondys 51, including improved motor function, has not been established. In making this decision, the FDA considered the potential risks associated with the drug, the life-threatening and debilitating nature of the disease for these children and the lack of available therapy.

Under the accelerated approval provisions, the FDA is requiring Sarepta Therapeutics to conduct a clinical trial to confirm the drug's clinical benefit. The required study is designed to assess whether Exondys 51 improves motor function of DMD patients with a confirmed mutation of the dystrophin gene amenable to exon 51 skipping. If the trial fails to verify clinical benefit, the FDA may initiate proceedings to withdraw approval of the drug.

The most common side effects reported by participants taking Exondys 51 in the clinical trials were balance disorder and vomiting.

The FDA granted Exondys 51 fast track designation (/patients/fast-track-breakthrough-therapy-accelerated-approval-and-priority-review/fast-track), which is a designation to facilitate the development and expedite the review of drugs that are intended to treat serious conditions and that demonstrate the potential to address an unmet medical need. It was also granted priority review and orphan drug designation. Priority review (/patients/fast-track-breakthrough-therapy-accelerated-approval-and-priority-review/priority-review) status is granted to applications for drugs that, if approved, would be a significant improvement in safety or effectiveness in the treatment of a serious condition. Orphan drug designation (/developing-products-rare-diseases-conditions) provides incentives such as clinical trial tax credits, user fee waiver and eligibility for orphan drug exclusivity to assist and encourage the development of drugs for rare diseases.

The manufacturer received a rare pediatric disease priority review voucher, which comes from a program intended to encourage development of new drugs and biologics for the prevention and treatment of rare pediatric diseases. This is the seventh rare pediatric disease priority review voucher issued by the FDA since the program began.

Exondys 51 is made by Sarepta Therapeutics of Cambridge, Massachusetts.

The FDA, an agency within the U.S. Department of Health and Human Services, protects the public health by assuring the safety, effectiveness, and security of human and veterinary drugs, vaccines and other biological products for human use, and medical devices. The agency also is responsible for the safety and security of our nation's food supply, cosmetics, dietary supplements, products that give off electronic radiation, and for regulating tobacco products.

Related Information

- National Institute of Neurological Disorders and Stroke: Muscular Dystrophy Information (http://www.ninds.nih.gov/disorders/md/md.htm)
- Centers for Disease Control and Prevention: Muscular Dystrophy (http://www.cdc.gov/ncbddd/musculardystrophy/)
- FDA Approved Drugs: Questions and Answers (/drugs/information-consumers-drugs/approved-drugs-questions-and-answers)
- FDA: New Drugs at FDA (/drug-innovation)
- FDA: Search Drugs@FDA (http://www.accessdata.fda.gov/scripts/cder/daf/index.cfm)

###

Inquiries

Media:

Sandy Walsh (mailto:sandy.walsh@fda.hhs.gov)

301-796-4669

Consumer:

888-INFO-FDA

More Press Announcements (/news-events/newsroom/press-announcements)